

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
15 November 2001 (15.11.2001)

PCT

(10) International Publication Number
WO 01/85791 A1

(51) International Patent Classification⁷: **C07K 14/705**, (74) Agents: **HINSCH, Matthew, E. et al.**; Townsend and Townsend and Crew LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111-3834 (US).

(21) International Application Number: **PCT/US01/15332**

(22) International Filing Date: 11 May 2001 (11.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/203,217 11 May 2000 (11.05.2000) US
60/205,945 18 May 2000 (18.05.2000) US

(71) Applicant (for all designated States except US): **LIFESPAN BIOSCIENCES, INC.** [US/US]; 700 Blanchard Street, Seattle, WA 98121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BROWN, Joseph, P.** [GB/US]; 411 West Prospect Street, Seattle, WA 98119 (US). **MILLER, Margaret** [US/US]; 7155 S.E. Maker Street, Mercer Island, WA 98040 (US). **BURMER, Glenna** [US/US]; 7516 55th Place, NE, Seattle, WA 98115 (US). **FABRE-SUVER, Christine** [FR/US]; 1212 NE 63rd Street, Seattle, WA 98115 (US). **PRITCHARD, David** [US/US]; 3429 Burke Avenue, N., Seattle, WA 98103 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A1

WO 01/85791

(54) Title: NUCLEIC ACID SEQUENCES FOR NOVEL GPCRS

(57) Abstract: The present invention is directed to new galanin receptors that are useful for treating and diagnosing a number of diseases and disorders, including, but not limited to, Alzheimer's disease, learning and memory disorders, hormonal problems, fat metabolism disorders, feeding disorders, pain perception disorders, diabetes, depression, etc. The present invention also provides methods for identifying modulators of galanin signaling. Such modulators are useful for treating the above-listed and other diseases and disorders.

BEST AVAILABLE COPY

NUCLEIC ACID SEQUENCES FOR NOVEL GPCRs

BACKGROUND OF THE INVENTION

Many physiologically important events are mediated by the binding of
5 guanine nucleotide-binding regulatory proteins (G proteins) to G protein-coupled
receptors (GPCRs). These events include vasodilation, stimulation or decrease in heart
rate, bronchodilation, stimulation of endocrine secretions and enhancement of gut
peristalsis, development, mitogenesis, cell proliferation and oncogenesis.

Guanine nucleotide-binding proteins are a family of proteins that transduce
10 signals from numerous cell surface receptors to downstream intracellular effector
molecules. G proteins are typically heterotrimeric proteins consisting of a guanyl-
nucleotide binding alpha subunit, a beta and a gamma subunits, the latter two being
tightly associated under physiological conditions (for a review, see, e.g., Conklin *et al.*,
Cell 73:631-641 (1993)). Each subunit is encoded by a separate gene. G proteins
15 commonly cycle between two forms, depending on whether GDP or GTP is bound to the
alpha subunit. Upon binding of a ligand to a G protein-coupled receptor, the GDP
molecule bound to the alpha subunit is exchanged for a GTP molecule resulting in the
dissociation of the α subunit from the β and γ subunits. The free alpha subunit and the
beta-gamma complex are capable of transmitting a signal to downstream elements of a
20 variety of signal transduction pathways, for example by binding to and activating adenylyl
cyclase. This fundamental scheme of events forms the basis for a multiplicity of different
cell signaling phenomena.

The different members of the G protein coupled receptors super-family
share a number of functional and structural characteristics. In particular, as described
25 above, GPCRs have the ability to stimulate the exchange of bound GDP for GTP on
associated G proteins alpha subunits in response to agonist binding. Structurally, GPCRs
typically contain seven hydrophobic transmembrane segments that are suggested to be
transmembrane helices of 20-30 amino acids connected by extracellular or cytoplasmic
loops (see, e.g., Kobilka *et al.*, *Science* 240:1310 (1988); Maggio *et al.*, *FEBS Lett.*
30 319:195 (1993); Maggio *et al.*, *Proc. Natl. Acad. Sci USA* 90:3103 (1993); Ridge *et al.*,
Proc. Natl. Sci USA 91:3204 (1995); Schonenberg *et al.*, *J. Biol. Chem.* 270:18000
(1995); Huang *et al.*, *J. Biol. Chem.* 256:3802 (1981); Popot *et al.*, *J. Mol. Biol.* 198:655

(1987); Kahn and Engelman, *Biochemistry* 31:6144 (1992); Schoneberg *et al.*, *EMBO J.* 15:1283 (1996); Wong *et al.*, *J. Biol. Chem.* 265:6219 (1990); Monnot *et al.*, *J. Biol. Chem.* 271:1507 (1996); Gudermann *et al.*, *Annu. Rev. Neurosci.* 20:399 (1997); Osuga *et al.*, *J. Biol. Chem.* 272:25006 (1997); Lefkowitz *et al.*, *J. Biol. Chem.* 263:4993-4996
5 (1988); Panayotou and Waterfield, *Curr. Opinion Cell Biol.* 1:167-176 (1989); and G Protein-Coupled Receptor Database, <http://www.gcrdb.uthscsa.edu>). In addition to G proteins, many enzymes, such as, for example, adenylate cyclase, cGMP phosphodiesterase and phospholipase C, can act as effectors for GPCRs' signal transduction (*see, e.g.*, Kinnamon & Margolskee, *Curr. Opin. Neurobiol.* 6:506-513
10 (1996)).

A large variety of molecules have been shown to be ligands for GPCRs. Identified ligands include, for example, purines, nucleotides and melatonin (*e.g.*, adenosine, cAMP, NTPs, *etc.*), biogenic amines (*e.g.*, adrenaline, dopamine, histamine, acetylcholine, noradrenaline, serotonin, *etc.*), peptides (*e.g.*, angiotensin, calcitonin, chemokine, Corticotropin Releasing Factor, galanin, Growth Hormone Releasing Hormone, Gastric Inhibitory Peptide, Glucagon, Neuropeptide Y, Neurotensin, Opoiod, Thrombin, Secretin, Somatostatin, Thyrotropin Releasing Hormone, Vasopressin, Vasoactive Intestinal Peptide, *etc.*), lipids and lipid-based compounds (*e.g.*, cannabinoids, Platelet Activating Factor, *etc.*), excitatory amino acids and ions (*e.g.*, glutamate, calcium, GABA, *etc.*), toxins, *etc.* In addition, there are many "orphan" G protein-coupled receptors (*e.g.*, some olfactory G protein-coupled receptors) for which ligands have not been identified.

G protein-coupled receptors thus play a central role in transducing numerous signals and regulating cellular metabolism. Accordingly, GPCRs have been implicated in a large number of diseases, such as, Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, a number of sarcomas (*e.g.*, chondrosarcoma, Ewing's sarcoma, osteosarcoma, *etc.*) and carcinomas (*e.g.*, basal cell carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, thyroid carcinoma, *etc.*), psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma multiform, Hodgkin's disease,
25
30

lymphoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, etc.

While many GPCRs have been identified, many more remain to be discovered. In addition, the specific GPCRs involved in the different biological processes, and in particular diseases, are not known.

Galanin is a widely distributed 28 amino acid peptide hormone which has been shown to regulate a variety of biological processes, including, for example, hormone release, neurotransmitter release, nociception, feeding behavior, cognitive function and reproductive behavior.

Galanin signaling has been shown to modulate the release of a variety of neurotransmitters, including, but not limited to, acetylcholine, norepinephrine, serotonin and dopamine (*see, e.g.*, Bartfai *Crit. Rev. Neurobiol.* 7:229 (1993)). Cumulative evidence suggests that galanin acts as an inhibitory cosecreted peptide. Galanin has been postulated to impair secretion of neurotransmitters by acting at the pre-synaptic autoreceptors as well as at the post-synaptic action site of these neurotransmitters. In particular, galanin inhibits acetylcholine release into the ventral hippocampus. Galanin may thus impair memory and learning by inhibiting the cholinergic function.

Galanin is to date the only neurotransmitter that has been shown to be upregulated in Alzheimer's disease. In addition, a variety of experiments, including the central injection of galanin and the generation of transgenic mice, have shown that the overexpression and/or oversecretion of galanin impairs performance of memory and learning tasks. These results indicate that the hypertrophy of galanin pathways contributes to the cognitive deficits in Alzheimer's disease.

Galanin has further been shown to inhibit the release of vasopressin and insulin, while it stimulates the release of growth hormone, prolactin and luteinizing hormone. Galanin has been shown to play a role in the control of fat metabolism, and body adiposity, which may be mediated by its effect on insulin. Galanin inhibits insulin secretion and, conversely, insulin injection inhibits central galanin expression. Galanin acts within the medial preoptic area and paraventricular nucleus to modulate fat intake and fat metabolism, but the specific subtype of galanin receptors involved in this function are not known. Galanin also acts within the supraoptic nucleus and paraventricular nucleus to modulate fluid balance. In addition, galanin regulates feeding behavior.

Galanin may exert neurotrophic and/or neuroprotective actions within the central nervous system. Treatment of rats with galanin has been shown to reduce

behavioral impairments following brain injury. Galanin gene expression is upregulated in injured neurons and this may contribute to cell survival. Despite the substantial loss of cells within the locus ceruleus, the percentage of noradrenergic neurons that coexpress galanin mRNA is increased in Alzheimer's disease supporting the idea that galanin may exert a neuroprotective effect.

5 Galanin is co-localized with gonadotropin-releasing hormone (GnRH) in the medial preoptic region of several species. The pattern of coexpression exhibits sexual dimorphism in rats. In both rats and monkeys, gonadal hormones regulate galanin expression in GnRH cells. Galanin, acting within the anterior pituitary, plays a role in the 10 regulation of luteinizing hormone release. Galanin facilitates sex behavior via actions within the medial preoptic regions.

Under normal conditions, galanin has potent antinociceptive effects. After peripheral nerve injury the inhibitory control exerted by endogenous galanin is increased. During inflammation, galanin expression within the dorsal horn is increased.
15 Endogenous galanin appears to play an enhanced antinociceptive role in chronic pain or neuropathic or inflammatory origin.

Galanin has been indicated in the etiology of depression. Galanin is colocalized within the serotonergic and noradrenergic systems. An increase in the amount of galanin released from ascending noradrenergic neurons into the ventral 20 tegmental area has been proposed to decrease dopamine release and thereby decrease motor activation and anhedonia, two major symptoms of depression. The receptors involved in these functions are not known.

Galanin has also been shown to control gastrointestinal and cardiovascular actions. For example, in the guinea pig ileum, galanin administration inhibits neurally 25 induced smooth muscle contractility probably via its ability to reduce acetylcholine release. In addition, galanin inhibits somatostatin and gastrin release. Galanin also decreases blood flow following injection into the mesenteric arteriole, as well as sodium and chloride net absorption.

Galanin thus plays an important role in a large variety of physiological 30 processes.

The effects of galanin are mediated via G-protein coupled receptors for which three types have been cloned, GALR1, GALR2 and GALR3 (see, e.g., Howard *et al.*, *FEBS letter*, 405:285-290 (1997); Bloomquist *et al.*, *Biochem. Biophys. Res. Commun.* 243:474-479 (1998); WO 98/15570; WO 99/31130; WO 97/46681; WO

97/26853). For most of the biological processes regulated by galanin, the specific receptors involved in these functions are not known.

Identifying additional G protein-coupled receptors would allow insight into the role of each receptor in the different biological processes in which GPCR-mediated signaling is involved. There is a strong need in the art for diagnostic and therapeutic tools for detection and treatment of the numerous diseases and disorders involving GPCR-mediated signaling. In addition, identifying additional receptors for galanin would allow insight into the role of each receptor in the different biological processes in which galanin is involved. Moreover, there is a strong need in the art for diagnostic and therapeutic tools for detection and treatment of the numerous diseases and disorders involving galanin signaling. This invention addresses these and other needs.

SUMMARY OF THE INVENTION

The present invention provides polypeptides having at least 70%, 75%, 15 80%, 85%, 90%, 95% or more identity with the polypeptides encoded by the nucleic acid molecules having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. In one embodiment, the polypeptides of the invention are encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. In other embodiments, the polypeptides 20 of the present invention comprise a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85%, and most preferably 90% or more, identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. In 25 some embodiments, the nucleic acids molecules encoding the polypeptides of the invention are operably linked to a heterologous promoter. The present invention also provides expression vectors comprising the nucleic acid molecules encoding the polypeptides of the invention, as well as host cells comprising the expression vectors. In one embodiment, the host cell is a mammalian cell.

30 The present invention is also directed to nucleic acid probes that specifically hybridize with the nucleic acid molecules encoding the described polypeptides. The probes can be DNA or RNA. Antisense nucleic acid molecules that specifically hybridize to the nucleic acid sequences encoding the polypeptides of the invention are also provided.

In another aspect, antibodies that specifically bind to the polypeptides of the invention are also provided. The antibodies can be monoclonal or polyclonal.

The antibodies and nucleic acid probes described above can be used to detect the presence of the polypeptides of the invention or of the nucleic acid molecules encoding the described polypeptides. They can be used to diagnose a variety of diseases and disorders in which G protein-coupled receptors are involved, such as, e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, etc.

The present invention is also directed to methods for identifying compounds that modulate the expression of one or more polypeptides of the invention, the methods comprising culturing a cell in the presence of a modulator to form a first cell culture, contacting RNA or cDNA from the first cell culture with at least one probe, each probe comprising a polynucleotide sequence encoding a polypeptide of the invention, and determining whether the amount of the probe(s) which hybridizes to the RNA or cDNA from the first cell culture is increased or decreased relative to the amount of the probe(s) which hybridizes to RNA or cDNA from a second cell culture grown in the absence of the modulator.

In addition, the present invention provides methods for identifying compounds that modulate the activity of one or more polypeptides of the invention, the methods comprising culturing cells expressing at least one polypeptide of interest in the presence of a compound, measuring the activity of the polypeptide(s) or second messenger activity and determining whether the activity is increased or decreased relative to the activity of the polypeptide(s) or second messenger activity from a second cell culture grown in the absence of the modulator.

The compounds identified using the methods of the present invention can be modulators, activators, repressors, agonists or antagonists and have therapeutic uses

for treating a variety of disorders and/or diseases in which G protein-coupled receptors have been implicated, such as, e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy,

5 embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis,

10 rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, etc.

The present invention provides is directed to polypeptides having at least 80% identity, optionally at least 85% identity, with the polypeptide encoded by the nucleic acid molecule having the nucleotide sequence set forth in SEQ ID NO:1. In one embodiment, the polypeptide of the present invention is the polypeptide encoded by the sequence set forth in SEQ ID NO:1. In other embodiments, the polypeptides of the present invention comprise a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85% and most preferably 90% or more identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from the polypeptide encoded by the nucleic acid molecule having the nucleotide sequence set forth in SEQ ID NO:1. Vectors comprising the nucleic acids encoding the polypeptides of the invention, and host cells comprising the expression vectors are also provided. In some embodiments, the nucleic acid molecules encoding the polypeptides of the invention are operably linked to a heterologous promoter. In some embodiments, the host cell is a mammalian cell.

The present invention is also directed to nucleic acid probes that specifically hybridize with the nucleic acid molecules encoding the polypeptides of the invention. The probes can be DNA or RNA. Antisense nucleic acid molecules that specifically hybridize to the nucleic acid molecules encoding the polypeptides of the invention are also provided.

In another aspect, antibodies that specifically bind to the polypeptides of the invention are also provided. The antibodies can be monoclonal or polyclonal.

The nucleic acid probes and antibodies described above can be used to detect the presence of the nucleic acid molecules encoding the polypeptides of the invention. They can be used to diagnose a variety of diseases and disorders in which galanin is involved, such as, cognition and memory disorders, anorexia, hormonal release 5 disorders, cardiovascular activity disorders, pain perception disorders, obesity, diabetes, Alzheimer's disease, etc.

The present invention is also directed to methods for identifying compounds that modulate the expression of the polypeptides of the invention, comprising culturing a cell in the presence of a modulator to form a first cell culture, contacting RNA or cDNA from the first cell culture with a probe which comprises a polynucleotide sequence encoding the polypeptide of the invention, and determining whether the amount 10 of the probe which hybridizes to the RNA or cDNA from the first cell culture is increased or decreased relative to the amount of the probe which hybridizes to RNA or cDNA from a second cell culture grown in the absence of the modulator.

In addition, the present invention provides a method for identifying compounds that modulate the activity of the polypeptides of the invention, comprising culturing cells expressing the polypeptide of interest in the presence of a compound, measuring the activity of the polypeptide or second messenger activity and determining whether the activity is increased or decreased relative to the activity of the polypeptide or 20 second messenger activity from a second cell culture grown in the absence of the modulator.

The compounds identified using the methods of the present invention can be modulators, activators, repressors, agonists or antagonists and have therapeutic uses for treating a variety of disorders and/or diseases in which galanin has been implicated. 25 For example, compounds that decrease the expression (repressors) or activity (antagonists) of the polypeptides of the invention can be used, e.g., to treat obesity, diabetes, hyperlipidemia, stroke, cognitive disorders, Alzheimer's disease, and/or endocrine disorders. Compounds that increase expression (activators) or activity (agonists) of the polypeptides of the invention can be used, for example, to treat anorexia 30 and to decrease noninception.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

I. INTRODUCTION

The present invention is directed to novel G protein-coupled receptors (GPCRs) that are useful for treating and diagnosing a number of diseases and disorders, including, but not limited to, Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy,

5 chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma,

10 pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, etc. The present invention also provides methods for identifying modulators of G protein-coupled receptor-mediated signaling. Such

15 modulators are useful for treating the above-listed and other diseases and disorders.

In some aspects, the present invention is directed to new galanin receptors that are useful for treating and diagnosing a number of diseases and disorders, including, but not limited to, Alzheimer's disease, learning and memory disorders, hormonal problems, fat metabolism disorders, feeding disorders, pain perception disorders,

20 diabetes, depression, etc. The present invention also provides methods for identifying modulators of galanin signaling. Such modulators are useful for treating the above-listed and other diseases and disorders.

The invention provides novel G protein-coupled receptors, as well as vectors and cells to express these novel GPCRs, including, e.g., galanin receptors. Probes

25 and antibodies that can be used to detect the GPCRs of the invention are also provided, as well as antisense polynucleotides. The probes and antibodies are useful for diagnostic purposes. In addition, the nucleic acids encoding the polypeptides of the invention, antisense polynucleotides and polypeptides of the invention are useful for gene therapy applications. The present invention also provides nucleic acid molecules encoding the

30 polypeptides of the invention operably linked to a heterologous promoter that drives expression of the protein encoded by the nucleic acid sequence.

The invention further provides methods of screening for modulators, e.g., activators, inhibitors, stimulators, enhancers, agonists, and antagonists, of these novel G protein-coupled receptors. Such modulators of the activity of the GPCRs are useful for

pharmacological and genetic modulation of the signaling pathways in which GPCRs are involved. These methods of screening can be used to identify high affinity agonists and antagonists of GPCRs' activity. These modulatory compounds can then be used in pharmaceutical industry to regulate G protein-coupled receptor-mediated signaling to treat a variety of diseases or disorders. Thus, the invention provides assays for GPCR-mediated signaling modulation, where the G protein-coupled receptors of the invention or other molecules located downstream of the G protein coupled receptor act as direct or indirect reporter molecules for the effect of modulators on GPCR-mediated signaling. G protein-coupled receptors can be used in assays, *e.g.*, to measure changes in ligand binding, transcription, signal transduction, receptor-ligand interactions, second messenger concentrations, *in vitro*, *in vivo*, and *ex vivo*.

In some embodiments, the present invention provides novel galanin receptors (GAL4), as well as vectors and cells to express the galanin receptors. Probes and antibodies that can be used to detect the galanin receptors of the invention are also provided, as well as antisense polynucleotides. The probes and antibodies are useful for diagnostic purposes. In addition, the nucleic acids encoding the polypeptides of the invention, antisense polynucleotides and polypeptides of the invention are useful for gene therapy applications.

In some aspects, the invention further provides methods of screening for modulators, *e.g.*, activators, inhibitors, stimulators, enhancers, agonists, and antagonists, of these novel galanin receptors. Such modulators of the activity of the galanin receptors are useful for pharmacological and genetic modulation of the galanin signaling pathways. These methods of screening can be used to identify high affinity agonists and antagonists of galanin receptors' activity. These modulatory compounds can then be used in pharmaceutical industry to regulate galanin signaling to treat a variety of diseases or disorders. Thus, the invention provides assays for galanin signaling modulation, where the galanin receptors of the invention or other molecules located downstream in the galanin signaling pathway act as direct or indirect reporter molecules for the effect of modulators on galanin signaling. Galanin receptors can be used in assays, *e.g.*, to measure changes in ligand binding, transcription, signal transduction, receptor-ligand interactions, second messenger concentrations, *in vitro*, *in vivo*, and *ex vivo*.

II. DEFINITIONS

"Amplification primers" are oligonucleotides comprising either natural or analog nucleotides that can serve as the basis for the amplification of a selected nucleic acid sequence. They include, for example, both polymerase chain reaction primers and ligase chain reaction oligonucleotides.

5 "Antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either
10 kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each
15 pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, e.g., as intact immunoglobulins or as a number of well
20 characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer.
25 The Fab' monomer is essentially an Fab with part of the hinge region (see, Paul (Ed.) *Fundamental Immunology*, Third Edition, Raven Press, NY (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also
30 includes antibody fragments either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv).

"Biological samples" refers to any tissue or liquid sample having genomic DNA or other nucleic acids (e.g., mRNA) or proteins. It refers to samples of cells or tissue from a normal healthy individual as well as samples of cells or tissue from a subject

suspected of having, *e.g.*, Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, a sarcoma (*e.g.*, chondrosarcoma, Ewing's sarcoma, osteosarcoma, *etc.*), a carcinoma (*e.g.*, basal cell 5 carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, thyroid carcinoma, *etc.*), psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma multiform, Hodgkin's disease, lymphoma, macular degeneration, 10 malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, or any other disease or disorder in which G protein-coupled receptors are involved, as well as learning and/or memory disorders, diabetes, pain perception disorders, anorexia, obesity, hormonal release problems, or any other disease or disorder in which galanin is involved..

15 The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

The term "isolated," when applied to a nucleic acid or protein, denotes that 20 the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present 25 in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames which flank the gene and encode a protein other than the gene of interest. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% 30 pure.

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides which have similar binding properties as the reference nucleic acid and are

metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon 5 substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.* (1992); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA 10 encoded by a gene.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino 15 acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner 20 similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and 25 an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is 30 different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, "conservatively modified variants" refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide.

Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 2) Aspartic acid (D), Glutamic acid (E);
- 30 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 7) Serine (S), Threonine (T); and

8) Cysteine (C), Methionine (M)
(*see, e.g., Creighton, Proteins (1984)*).

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization,
5 *see, e.g., Alberts et al., Molecular Biology of the Cell (3rd ed., 1994) and Cantor and Schimmel, Biophysical Chemistry Part I: The Conformation of Biological Macromolecules (1980)*. "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains.
10 Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically 50 to 350 amino acids long. Typical domains are made up of sections of lesser organization such as stretches of β -sheet and α -helices. "Tertiary structure" refers to the complete three dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three dimensional structure formed by the noncovalent association
15 of independent tertiary units. Anisotropic terms are also known as energy terms.

"Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or
20 deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

25 The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one
30 of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence. Optionally, the identity exists over a

region that is at least about 50 amino acids or nucleotides in length, or more preferably over a region that is 75-100 amino acids or nucleotides in length.

The term "similarity," or percent "similarity," in the context of two or more polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of amino acid residues that are either the same or similar as defined in the 8 conservative amino acid substitutions defined above (*i.e.*, 60%, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% similar over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially similar." Optionally, this identity exists over a region that is at least about 50 amino acids in length, or more preferably over a region that is at least about 75-100 amino acids in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575

Science Dr., Madison, WI), or by manual alignment and visual inspection (see, e.g., Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle (1987) *J. Mol. Evol.* 35:351-360. The method used is similar to the method described by Higgins and Sharp (1989) *CABIOS* 5:151-153. The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, e.g., version 7.0 (Devereaux *et al.* (1984) *Nuc. Acids Res.* 12:387-395).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment

score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

5 The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in

10 10 Tijssen, *Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at

15 15 which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about

20 20 30° C for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as

25 25 following: 50% formamide, 5X SSC, and 1% SDS, incubating at 42° C, or 5X SSC, 1% SDS, incubating at 65° C, with wash in 0.2X SSC, and 0.1% SDS at 65° C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37° C, and a wash in 1X SSC at 45° C. Such

washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

5 For PCR, a temperature of about 36°C is typical for low stringency amplification, although annealing temperatures may vary between about 32°C and 48°C depending on primer length. For high stringency PCR amplification, a temperature of about 62°C is typical, although high stringency annealing temperatures can range from about 50°C to about 65°C, depending on the primer length and specificity. Typical cycle 10 conditions for both high and low stringency amplifications include a denaturation phase of 90°C - 95°C for 30 sec - 2 min., an annealing phase lasting 30 sec. - 2 min., and an extension phase of about 72°C for 1 - 2 min.

As used herein a "nucleic acid probe" is defined as a nucleic acid capable of binding to a target nucleic acid (*e.g.*, a nucleic acid encoding a galanin receptor) of 15 complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (*i.e.*, A, G, C, or T) or modified bases (7-deazaguanosine, inosine, *etc.*). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for 20 example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions.

Nucleic acid probes can be DNA or RNA fragments. DNA fragments can 25 be prepared, for example, by digesting plasmid DNA, or by use of PCR, or synthesized by either the phosphoramidite method described by Beaucage and Carruthers

A "labeled nucleic acid probe" is a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen bonds to a label such that the presence of the probe may be determined by detecting the presence of the label bound to the probe.

5 The phrase "a nucleic acid sequence encoding" refers to a nucleic acid which contains sequence information for a structural RNA such as rRNA, a tRNA, or the primary amino acid sequence of a specific protein or peptide, or a binding site for a trans-acting regulatory agent. This phrase specifically encompasses degenerate codons (*i.e.*, different codons which encode a single amino acid) of the native sequence or sequences
10 which may be introduced to conform with codon preference in a specific host cell.

The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so
15 modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not
20 found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to
25 each other in nature (*e.g.*, a fusion protein).

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes
30 distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation.
The term "operably linked" refers to a functional linkage between a nucleic acid

expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

The phrase "specifically (or selectively) binds to an antibody" or "specifically (or selectively) immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample.

Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised against a protein having an amino acid sequence encoded by any of the polynucleotides of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins, except for polymorphic variants. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. *See, Harlow and Lane Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, NY (1988)* for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically, a specific or selective reaction will be at least twice the background signal or noise and more typically more than 10 to 100 times background.

"Inhibitors," "activators," and "modulators" of G protein-coupled receptors expression or of G protein-coupled receptors' activity are used to refer to inhibitory, activating, or modulating molecules, respectively, identified using *in vitro* and *in vivo* assays for G protein-coupled receptors expression or G protein-mediated signaling, *e.g.,* ligands, agonists, antagonists, and their homologs and mimetics. Inhibitors are compounds that, *e.g.,* inhibit expression of a G protein-coupled receptor or bind to, partially or totally block stimulation, decrease, prevent, delay activation,

inactivate, desensitize, or down-regulate the activity of a G protein-coupled receptor, *e.g.*, antagonists. Activators are compounds that, *e.g.*, induce or activate the expression of a G protein-coupled receptor or bind to, stimulate, increase, open, activate, facilitate, enhance activation, sensitize or up-regulate the activity of G protein-coupled receptors, *e.g.*,

5 agonists. Modulators include compounds that, *e.g.*, alter the interaction of a receptor with extracellular proteins that bind activators or inhibitors, G proteins, and kinases. Modulators include genetically modified versions of G protein-coupled receptors, *e.g.*, with altered activity, as well as naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Assays for inhibitors, activators and

10 modulators include, *e.g.*, expressing a G protein-coupled receptor in cells or cell membranes, applying putative modulator compounds, in the presence or absence of a GPCR ligand (such as galanin, where appropriate) and then determining the functional effects on G protein-mediated signaling, as described above. Samples or assays comprising G protein-coupled receptors that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of inhibition. Control samples (untreated with inhibitors) are assigned a relative G protein-coupled receptor activity value of 100%. Inhibition of a G protein-coupled receptor is achieved when the G protein-coupled receptor activity value relative to the control is about 80%, optionally 50% or 25-0%.

15 Activation of a G protein-coupled receptor is achieved when the G protein-coupled receptor activity value relative to the control is 110%, optionally 150%, optionally 200-500%, or 1000-3000% higher.

III. GENERAL RECOMBINANT NUCLEIC ACIDS METHODS FOR USE WITH THE INVENTION

25 In numerous embodiments of the present invention, nucleic acids encoding the GPCRs of interest will be isolated and cloned using recombinant methods. Such embodiments are used, *e.g.*, to isolate GPCR-encoding polynucleotides for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from GPCRs, to monitor GPCR gene expression, for the isolation or

30 detection of GPCR sequences in different species, for diagnostic purposes in a patient, *e.g.*, to detect mutations in GPCRs, *etc.* In one embodiment, the nucleic acids of the invention are from any mammal, including, in particular, *e.g.*, a human, a rat, a mouse, *etc.*

In addition, recombinant expression of a GPCR of interest in eukaryotic cells, is useful for making cell membrane preparations that can be used for receptor binding assays. Receptor binding assays are used, in particular, for screening for modulators of the activity of GPCRs.

5 **A. General Recombinant Nucleic Acids Methods**

The numerous applications of the present invention involving the cloning, synthesis, maintenance, mutagenesis, and other manipulations of nucleic acid sequences can be performed using routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*,
10 *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and Ausubel *et al.*, *Current Protocols in Molecular Biology* (1994).

Nucleotide sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis or,
15 alternatively, from published DNA sequences.

Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage and Caruthers, *Tetrahedron Letts.* 22(20):1859-1862 (1981), using an automated synthesizer, as described in Needham Van Devanter *et al.*, *Nucleic Acids Res.*
20 12:6159-6168 (1984). Purification of oligonucleotides is, for example, by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson and Reanier, *J. Chrom.* 255:137-149 (1983).

The nucleic acids described here, or fragments thereof, can be used as hybridization probes for genomic or cDNA libraries to isolate the corresponding complete
25 gene (including regulatory and promoter regions, exons and introns) or cDNAs, in particular cDNA clones corresponding to full-length transcripts. The probes may also be used to isolate other genes and cDNAs which have a high sequence similarity to the gene of interest or similar biological activity. Probes of this type preferably have at least 30 bases and may contain, for example, 50 or more bases.

30 The sequence of the cloned genes and synthetic oligonucleotides can be verified using the chemical degradation method of Maxam and Gilbert, *Methods in Enzymology* 65:499-560 (1980). The sequence can be confirmed after the assembly of the oligonucleotide fragments into the double-stranded DNA sequence using the method

of Maxam and Gilbert, *supra*, or the chain termination method for sequencing double-stranded templates of Wallace *et al.*, *Gene* 16:21-26 (1981). Southern blot hybridization techniques can be carried out according to Southern *et al.*, *J. Mol. Biol.* 98:503 (1975).

5 **B. Cloning Methods for the Isolation of Nucleotide Sequences Encoding the Desired Proteins**

In general, the nucleic acids encoding the subject proteins are cloned from DNA sequence libraries that are made to encode copy DNA (cDNA) or genomic DNA. The particular sequences can be located by hybridizing with an oligonucleotide probe, the sequence of which can be derived from the sequences provided herein (e.g., the sequences 10 set forth in Table 1), which provides a reference for PCR primers and defines suitable regions for isolating G protein-coupled receptors specific probes. Alternatively, where the sequence is cloned into an expression library, the expressed recombinant protein can be detected immunologically with antisera or purified antibodies made against the G protein-coupled receptor of interest.

15 Methods for making and screening genomic and cDNA libraries are well-known to those of skill in the art (see, e.g., Gubler and Hoffman, *Gene* 25:263-269 (1983); Benton and Davis, *Science* 196:180-182 (1977); and Sambrook, *supra*).

Briefly, to make the cDNA library, one should choose a source that is rich 20 in mRNA. The mRNA can then be made into cDNA, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. For a genomic library, the DNA is extracted from a suitable tissue and either mechanically sheared or enzymatically digested to yield fragments of preferably about 5-100 kb. The 25 fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*, and the recombinant phages are analyzed by plaque hybridization. Colony hybridization is carried out as generally described in Grunstein *et al.*, *Proc. Natl. Acad. Sci. USA* 72:3961-3965 (1975).

An alternative method combines the use of synthetic oligonucleotide 30 primers with polymerase extension on an mRNA or DNA template. Suitable primers can be designed from specific GPCRs, e.g., the sequences described in Table 1. This polymerase chain reaction (PCR) method amplifies the nucleic acids encoding the protein of interest directly from mRNA, cDNA, genomic libraries or cDNA libraries. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or

other *in vitro* amplification methods may also be useful, for example, to clone nucleic acids encoding specific proteins and express said proteins, to synthesize nucleic acids that will be used as probes for detecting the presence of mRNA encoding a G protein-coupled receptor of the invention in physiological samples, for nucleic acid sequencing, or for other purposes (see, U.S. Patent Nos. 4,683,195 and 4,683,202). Genes amplified by a PCR reaction can be purified, e.g., from agarose gels, and cloned into an appropriate vector.

Appropriate primers and probes for identifying the genes encoding the G protein-coupled receptors of the invention from mammalian tissues can be derived from the sequences provided herein, in particular the sequences set forth in Table 1. For a general overview of PCR, see, Innis *et al.*, *PCR Protocols: A Guide to Methods and Applications*, Academic Press, San Diego (1990).

Synthetic oligonucleotides can be used to construct genes. This is done using a series of overlapping oligonucleotides, usually 40-120 bp in length, representing both the sense and anti-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned.

A gene encoding a G protein-coupled receptor of the invention can be cloned using intermediate vectors before transformation into mammalian cells for expression. These intermediate vectors are typically prokaryote vectors or shuttle vectors. The proteins can be expressed in either prokaryotes, using standard methods well-known to those of skill in the art, or eukaryotes as described *infra*.

C. Expression in Eukaryotes

Standard eukaryotic transfection methods are used to produce eukaryotic cell lines, e.g., yeast, insect, or mammalian cell lines, which express large quantities of the G protein-coupled receptors of the invention which are then purified using standard techniques (see, e.g., Colley *et al.*, *J. Biol. Chem.* 264:17619-17622, (1989); and *Guide to Protein Purification*, in Vol. 182 of *Methods in Enzymology* (Deutscher ed., 1990)).

Transformations of eukaryotic cells are performed according to standard techniques as described by Morrison, *J. Bact.*, 132:349-351 (1977), or by Clark-Curtiss and Curtiss, *Methods in Enzymology*, 101:347-362 R. Wu *et al.* (Eds) Academic Press, NY (1983).

Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate

transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, plasma vectors, viral vectors and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see Sambrook et al., supra*). It is only necessary that the particular genetic 5 engineering procedure utilized be capable of successfully introducing at least one gene into the host cell which is capable of expressing the protein.

The particular eukaryotic expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic cells may be used. Expression vectors containing regulatory 10 elements from eukaryotic viruses are typically used. Suitable vectors for use in the present invention include, but are not limited to, SV40 vectors, vectors derived from bovine papilloma virus or from the Epstein Barr virus and baculovirus vectors, and any other vector allowing expression of proteins under the direction of the SV-40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous 15 sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

The vectors usually include selectable markers which result in gene amplification, such as, e.g., thymidine kinase, aminoglycoside phosphotransferase, hygromycin B phosphotransferase, xanthine-guanine phosphoribosyl transferase, CAD 20 (carbamyl phosphate synthetase, aspartate transcarbamylase, and dihydroorotase), adenosine deaminase, dihydrofolate reductase, asparagine synthetase and ouabain selection. Alternatively, high yield expression systems not involving gene amplification are also suitable, such as, e.g., using a baculovirus vector in insect cells, with a target 25 protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The expression vector of the present invention will typically contain both prokaryotic sequences that facilitate the cloning of the vector in bacteria as well as one or more eukaryotic transcription units that are expressed only in eukaryotic cells, such as mammalian cells. The vector may or may not comprise a eukaryotic replicon. If a 30 eukaryotic replicon is present, then the vector is amplifiable in eukaryotic cells using the appropriate selectable marker. If the vector does not comprise a eukaryotic replicon, no episomal amplification is possible. Instead, the transfected DNA integrates into the genome of the transfected cell, where the promoter directs expression of the desired gene. The expression vector is typically constructed from elements derived from different, well

characterized viral or mammalian genes. For a general discussion of the expression of cloned genes in cultured mammalian cells, *see, Sambrook et al., supra, Ch. 16.*

The prokaryotic elements that are typically included in the mammalian expression vector include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of eukaryotic sequences. The particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable. The prokaryotic sequences are preferably chosen such that they do not interfere with the replication of the DNA in eukaryotic cells.

The expression vector contains a eukaryotic transcription unit or expression cassette that contains all the elements required for the expression of the DNA encoding the G protein-coupled receptors of interest in eukaryotic cells. A typical expression cassette contains a promoter operably linked to the DNA sequence encoding the G protein-coupled receptor and signals required for efficient polyadenylation of the transcript. The DNA sequence encoding the protein may typically be linked to a cleavable signal peptide sequence to promote secretion of the encoded protein by the transformed cell. Such signal peptides would include, among others, the signal peptides from tissue plasminogen activator, insulin, and neuron growth factor, and juvenile hormone esterase of *Heliothis virescens*. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

Eukaryotic promoters typically contain two types of recognition sequences, the TATA box and upstream promoter elements. The TATA box, located 25-30 base pairs upstream of the transcription initiation site, is thought to be involved in directing RNA polymerase to begin RNA synthesis. The other upstream promoter elements determine the rate at which transcription is initiated.

Enhancer elements can stimulate transcription up to 1,000 fold from linked homologous or heterologous promoters. Enhancers are active when placed downstream or upstream from the transcription initiation site. Many enhancer elements derived from viruses have a broad host range and are active in a variety of tissues (*see, Enhancers and Eukaryotic Expression, Cold Spring Harbor Pres, Cold Spring Harbor, NY (1983)*).

In the construction of the expression cassette, the promoter is preferably positioned at about the same distance from the heterologous transcription start site as it is

from the transcription start site in its natural setting. As is known in the art, some variation in this distance can, however, be accommodated without loss of promoter function.

In addition to a promoter sequence, the expression cassette should also

5 contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from a different gene.

If the mRNA encoded by the structural gene is to be efficiently translated, polyadenylation sequences are also commonly added to the vector construct. Two

10 distinct sequence elements are required for accurate and efficient polyadenylation: GU or U rich sequences located downstream from the polyadenylation site and a highly conserved sequence of six nucleotides, AAUAAA, located 11-30 nucleotides upstream. Termination and polyadenylation signals that are suitable for the present invention include those derived from SV40, or a partial genomic copy of a gene already resident on

15 the expression vector.

In addition to the elements already described, the expression vector of the present invention may typically contain other specialized elements intended to increase the level of expression of cloned genes or to facilitate the identification of cells that carry the transfected DNA. For instance, a number of animal viruses contain DNA sequences

20 that promote the extra chromosomal replication of the viral genome in permissive cell types. Plasmids bearing these viral replicons are replicated episomally as long as the appropriate factors are provided by genes either carried on the plasmid or with the genome of the host cell.

The cDNA encoding the protein of interest can be ligated to various

25 expression vectors for use in transforming host cell cultures. The vectors typically contain gene sequences to initiate transcription and translation of the G protein-coupled receptor gene. These sequences need to be compatible with the selected host cell. In addition, the vectors preferably contain a marker to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or metallothionein.

30 Additionally, a vector might contain a replicative origin.

Cells of mammalian origin are illustrative of cell cultures useful for the production of, for example, a G protein-coupled receptor of interest. Mammalian cell systems often will be in the form of monolayers of cells, although mammalian cell suspensions may also be used. Illustrative examples of mammalian cell lines include

VERO and HeLa cells, NIH 3T3, COS, Chinese hamster ovary (CHO), WI38, BHK, COS-7 or MDCK cell lines.

As indicated above, the vector, *e.g.*, a plasmid, which is used to transform the host cell, preferably contains DNA sequences to initiate transcription and sequences 5 to control the translation of the gene sequence encoding the G protein-coupled receptor of interest. These sequences are referred to as expression control sequences. Illustrative expression control sequences are described, *e.g.*, in Berman *et al.*, *Science*, 222:524-527 (1983); Thomsen *et al.*, *Proc. Natl. Acad. Sci.* 81:659-663 (1984); and Brinster *et al.*, *Nature* 296:39-42 (1982). The cloning vector containing the expression control 10 sequences is cleaved using restriction enzymes, adjusted in size as necessary or desirable and ligated with sequences encoding the G protein-coupled receptor by means well-known in the art.

When higher animal host cells are employed, polyadenylation or transcription terminator sequences from known mammalian genes need to be incorporated 15 into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague *et al.*, *J. Virol.* 45:773-781 (1983)).

Additionally, gene sequences to control replication in the host cell may be 20 incorporated into the vector such as those found in bovine papilloma virus type-vectors (*see*, Saveria-Campo, "Bovine Papilloma virus DNA a Eukaryotic Cloning Vector" In: *DNA Cloning Vol.II: a Practical Approach* (Glover Ed.), IRL Press, Arlington, Virginia pp. 213-238 (1985)).

The transformed cells are cultured by means well-known in the art. For 25 example, such means are published in *Biochemical Methods in Cell Culture and Virology*, Kuchler, Dowden, Hutchinson and Ross, Inc. (1977). The expressed protein is isolated from cells grown as suspensions or as monolayers. The latter are recovered by well-known mechanical, chemical or enzymatic means.

IV. PURIFICATION OF THE PROTEINS FOR USE WITH THE INVENTION

30 After expression, the proteins of the present invention can be purified to substantial purity by standard techniques, including selective precipitation with substances as ammonium sulfate, column chromatography, immunopurification methods, and other methods known to those of skill in the art (*see, e.g.*, Scopes *Protein*

Purification: Principles and Practice, Springer-Verlag, NY (1982); U.S. Patent No. 4,673,641; Ausubel *et al.*, *supra*; and Sambrook *et al.*, *supra*).

A number of conventional procedures can be employed when a recombinant protein is being purified. For example, proteins having established 5 molecular adhesion properties can be reversibly fused to the subject protein. With the appropriate ligand, a G protein-coupled receptor of interest, for example, can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein is then removed by enzymatic activity. Finally, the G protein-coupled receptors of the invention can be purified using immunoaffinity 10 columns.

A. Purification of Proteins from Recombinant Bacteria

When recombinant proteins are expressed by the transformed bacteria in large amounts, typically after promoter induction, although expression can be constitutive, the proteins may form insoluble aggregates. There are several protocols that 15 are suitable for purification of protein inclusion bodies. For example, purification of aggregate proteins (hereinafter referred to as inclusion bodies) typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells typically, but not limited to, by incubation in a buffer of about 100-150 µg/ml lysozyme and 0.1% Nonidet P40, a non-ionic detergent. The cell suspension can be 20 ground using a Polytron grinder (Brinkman Instruments, Westbury, NY). Alternatively, the cells can be sonicated on ice. Alternate methods of lysing bacteria are described in Ausubel *et al.*, and Sambrook *et al.*, both *supra*, and will be apparent to those of skill in the art.

The cell suspension is generally centrifuged and the pellet containing the 25 inclusion bodies resuspended in buffer which does not dissolve but washes the inclusion bodies, e.g., 20 mM Tris-HCl (pH 7.2), 1 mM EDTA, 150 mM NaCl and 2% Triton-X 100, a non-ionic detergent. It may be necessary to repeat the wash step to remove as much cellular debris as possible. The remaining pellet of inclusion bodies may be resuspended in an appropriate buffer (e.g., 20 mM sodium phosphate, pH 6.8, 150 mM 30 NaCl). Other appropriate buffers will be apparent to those of skill in the art.

Following the washing step, the inclusion bodies are solubilized by the addition of a solvent that is both a strong hydrogen acceptor and a strong hydrogen donor (or a combination of solvents each having one of these properties). The proteins that

formed the inclusion bodies may then be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to, urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents which are capable of 5 solubilizing aggregate-forming proteins, such as SDS (sodium dodecyl sulfate) and 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, 10 this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of the immunologically and/or biologically active protein of interest. After solubilization, the protein can be separated from other bacterial proteins by standard separation techniques.

Alternatively, it is possible to purify proteins from bacteria periplasm. Where the protein is exported into the periplasm of the bacteria, the periplasmic fraction 15 of the bacteria can be isolated by cold osmotic shock in addition to other methods known to those of skill in the art (*see, Ausubel et al., supra*). To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM MgSO₄ and kept in an ice bath 20 for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well-known to those of skill in the art.

B. Standard Protein Separation Techniques For Purifying Proteins

25 1. Solubility Fractionation

Often as an initial step, and if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing 30 the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol is to add saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is

between 20-30%. This will precipitate the most hydrophobic proteins. The precipitate is discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, through 5 either dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well-known to those of skill in the art and can be used to fractionate complex protein mixtures.

2. Size Differential Filtration

Based on a calculated molecular weight, a protein of greater and lesser size 10 can be isolated using ultrafiltration through membranes of different pore sizes (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the 15 molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

3. Column Chromatography

The proteins of interest can also be separated from other proteins on the 20 basis of their size, net surface charge, hydrophobicity and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well-known in the art.

It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., 25 Pharmacia Biotech).

V. DETECTION OF GENE EXPRESSION OF THE GPCRs

The polypeptides of the present invention and the polynucleotides encoding them can be employed as research reagents and materials for discovery of treatments and diagnostics to human disease. It will be readily apparent to those of skill 30 in the art that although the following discussion is directed to methods for detecting nucleic acids encoding a G protein-coupled receptor, similar methods can be used to detect nucleic acids associated with, e.g., Alzheimer's disease, depression, specific carcinomas and sarcomas, or any disease or disorder in which GPCR-mediated signaling

is involved. In aspects involving, e.g., a galanin receptor, similar methods can be used to detect nucleic acids associated with, e.g., Alzheimer's disease, learning and memory disorders, reproduction and sex behavior disorders, feeding disorders, fat metabolism and body adiposity, regulation of neurotransmitter release, pain perception, depression, 5 regulation of hormone release, cardiovascular actions regulation, or any disease or disorder in which galanin signaling is involved.

As should be apparent to those of skill in the art, the invention is based, at least in part, in the identification of novel G protein-coupled receptors, including a novel galanin receptor (GAL4). Accordingly, the present invention also includes methods for 10 detecting the presence, alteration or absence of nucleic acids (e.g., DNA or RNA) encoding such G protein-coupled receptors in a physiological specimen in order to determine the presence of, e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, 15 embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, 20 rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, etc., associated with mutations created in the sequences encoding the GPCRs that modify the expression and/or activity of the receptors, including those disorders associated with mutations created in the sequences encoding the galanin 25 receptor that modify the activity of the receptor, including cognitive deficit, Alzheimer's disease, reproductive disorder, fat metabolism disorder, inhibition of neurotransmitter release, pain perception disorder, depression, hormone release disorder, decrease in blood flow, etc. Any tissue having cells bearing the genome of an individual, or RNA encoding the GPCRs can be used as well as biopsies of suspect tissue. It is also possible and 30 preferred in some circumstances to conduct assays on cells that are isolated under microscopic visualization. A particularly useful method is the microdissection technique described in WO 95/23960. The cells isolated by microscopic visualization can be used in any of the assays described herein including both genomic and immunological based assays.

This invention provides methods of genotyping family members in which relatives are diagnosed with, e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, 5 embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, 10 rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis, thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, Alzheimer's disease, depression, fat metabolism disorders, anorexia, stroke, diabetes, etc. Conventional methods of genotyping are known to those of skill in the art.

15 The probes are capable of binding to a target nucleic acid (e.g., a nucleic acid encoding a G protein-coupled receptor of interest). By assaying for the presence or absence of the probe, one can detect the presence or absence of the target nucleic acid in a sample. Preferably, non-hybridizing probe and target nucleic acids are removed (e.g., by washing) prior to detecting the presence of the probe.

20 A variety of methods of specific DNA and RNA measurement using nucleic acid hybridization techniques are known to those of skill in the art (see, Sambrook, *supra*). Some methods involve an electrophoretic separation (e.g., Southern blot for detecting DNA, and Northern blot for detecting RNA), but measurement of DNA and RNA can also be carried out in the absence of electrophoretic separation (e.g., by dot blot). Southern blot of genomic DNA (e.g., from a human) can be used for screening for 25 restriction fragment length polymorphism (RFLP) to detect the presence of a genetic disorder affecting a G protein-coupled receptor of the invention.

The selection of a nucleic acid hybridization format is not critical. A variety of nucleic acid hybridization formats are known to those skilled in the art. For 30 example, common formats include sandwich assays and competition or displacement assays. Hybridization techniques are generally described in Hames and Higgins, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press (1985); Gall and Pardue, *Proc. Natl. Acad. Sci. U.S.A.*, 63:378-383 (1969); and John *et al.*, *Nature*, 223:582-587 (1969).

Detection of a hybridization complex may require the binding of a signal generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal. The binding of the 5 signal generation complex is also readily amenable to accelerations by exposure to ultrasonic energy.

The label may also allow indirect detection of the hybridization complex. For example, where the label is a hapten or antigen, the sample can be detected by using antibodies. In these systems, a signal is generated by attaching fluorescent or enzyme 10 molecules to the antibodies or in some cases, by attachment to a radioactive label (see, e.g., Tijssen, "Practice and Theory of Enzyme Immunoassays," *Laboratory Techniques in Biochemistry and Molecular Biology*, pp. 9-20, Burdon and van Knippenberg Eds., Elsevier (1985)).

The probes are typically labeled either directly, as with isotopes, 15 chromophores, lumiphores, chromogens, or indirectly, such as with biotin, to which a streptavidin complex may later bind. Thus, the detectable labels used in the assays of the present invention can be primary labels (where the label comprises an element that is detected directly or that produces a directly detectable element) or secondary labels (where the detected label binds to a primary label, e.g., as is common in immunological 20 labeling). Typically, labeled signal nucleic acids are used to detect hybridization. Complementary nucleic acids or signal nucleic acids may be labeled by any one of several methods typically used to detect the presence of hybridized polynucleotides. The most common method of detection is the use of autoradiography with ^3H , ^{125}I , ^{35}S , ^{14}C , or ^{32}P -labeled probes or the like.

25 Other labels include, e.g., ligands which bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair members for a labeled ligand. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden, *Introduction to Immunocytochemistry*, 2nd ed., Springer Verlag, NY (1997); and in Haugland, *Handbook 30 of Fluorescent Probes and Research Chemicals*, a combined handbook and catalogue Published by Molecular Probes, Inc. (1996).

In general, a detector which monitors a particular probe or probe combination is used to detect the detection reagent label. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters,

cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill in the art. Commonly, an optical image of a substrate comprising bound labeling moieties is digitized for subsequent computer analysis.

5 Most typically, the amount of, for example, a G protein-coupled receptor RNA is measured by quantitating the amount of label fixed to the solid support by binding of the detection reagent. Typically, the presence of a modulator during incubation will increase or decrease the amount of label fixed to the solid support relative to a control incubation which does not comprise the modulator, or as compared to a
10 baseline established for a particular reaction type. Means of detecting and quantitating labels are well-known to those of skill in the art.

In preferred embodiments, the target nucleic acid or the probe is immobilized on a solid support. Solid supports suitable for use in the assays of the invention are known to those of skill in the art. As used herein, a solid support is a matrix
15 of material in a substantially fixed arrangement.

A variety of automated solid-phase assay techniques are also appropriate. For instance, very large scale immobilized polymer arrays (VLSIPS™), available from Affymetrix, Inc. in Santa Clara, CA, can be used to detect changes in expression levels of a plurality of genes involved in the same regulatory pathways simultaneously. See,
20 Tijssen, *supra.*, Fodor *et al.*, *Science*, 251:767-777 (1991); Sheldon *et al.*, *Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al.*, *Nature Medicine* 2(7):753-759 (1996). Thus, in one embodiment, the invention provides methods of detecting expression levels of the G protein-coupled receptors of the invention in combination with other G protein-coupled receptors and other nucleic acids known to be involved in regulating, e.g.,
25 Alzheimer's disease, depression, feeding behavior, diabetes, obesity, stroke, cognition and memory, hormone release, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, COPD, Crohn's disease, depression, Duchenne muscular dystrophy, embryonal carcinoma, epilepsy, Ewing's sarcoma, glioblastoma multiform, Hodgkin's disease, lymphoma, lung
30 adenocarcinoma, lung small cell carcinoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, multiple sclerosis, osteoarthritis, osteoporosis, osteosarcoma, ovarian carcinoma, pancreatic carcinoma, Parkinson's disease, prostate carcinoma, psoriasis, rhabdomyosarcoma, renal cell carcinoma, rheumatoid arthritis, schizophrenia, seminoma, squamous cell carcinoma, tuberculosis,

thyroid carcinoma, tonsil, transitional carcinoma of the bladder, ulcerative colitis, *etc.*, in which nucleic acids (*e.g.*, RNA from a cell culture) are hybridized to an array of nucleic acids that are known to be associated with the above-listed diseases and disorders. Thus, in one embodiment, the invention provides methods for detecting the expression levels of nucleic acids encoding the G protein-coupled receptors of the invention, in which nucleic acids (*e.g.*, RNA from a cell culture) are hybridized to an array of nucleic acids that are known to be associated with the above-listed diseases and disorders in which GPCRs have been implicated. In a second embodiment, the invention provides methods for detecting the expression levels of nucleic acids encoding the galanin receptors of the invention, in which nucleic acids (*e.g.*, RNA from a cell culture) are hybridized to an array of nucleic acids that are known to be associated with Alzheimer's disease, depression, fat metabolism disorders, feeding disorders, hormonal disorders, *etc.* For example, in the assay described *supra*, oligonucleotides which hybridize to a plurality of nucleic acids encoding either G protein-coupled receptors or other molecules known to be involved in the above-mentioned diseases and disorders are optionally synthesized on a DNA chip (such chips are available from Affymetrix) and the RNA from a biological sample, such as a cell culture, is hybridized to the chip for simultaneous analysis of multiple nucleic acids. The nucleic acids encoding the G protein-coupled receptors that are present in the sample which is assayed are detected at specific positions on the chip.

Detection can be accomplished, for example, by using a labeled detection moiety that binds specifically to duplex nucleic acids (*e.g.*, an antibody that is specific for RNA-DNA duplexes). One preferred example uses an antibody that recognizes DNA-RNA heteroduplexes in which the antibody is linked to an enzyme (typically by recombinant or covalent chemical bonding). The antibody is detected when the enzyme reacts with its substrate, producing a detectable product. Coutlee *et al.*, *Analytical Biochemistry* 181:153-162 (1989); Bogulavski *et al.*, *J. Immunol. Methods* 89:123-130 (1986); Prooijen-Knegt, *Exp. Cell Res.* 141:397-407 (1982); Rudkin, *Nature* 265:472-473 (1976); Stollar, *PNAS* 65:993-1000 (1970); Ballard, *Mol. Immunol.* 19:793-799 (1982); Pisetsky and Caster, *Mol. Immunol.* 19:645-650 (1982); Viscidi *et al.*, *J. Clin. Microbial.* 41:199-209 (1988); and Kiney *et al.*, *J. Clin. Microbiol.* 27:6-12 (1989) describe antibodies to RNA duplexes, including homo and heteroduplexes. Kits comprising antibodies specific for DNA:RNA hybrids are available, *e.g.*, from Digene Diagnostics, Inc. (Beltsville, MD).

In addition to available antibodies, one of skill in the art can easily make antibodies specific for nucleic acid duplexes using existing techniques, or modify those antibodies which are commercially or publicly available. In addition to the art referenced above, general methods for producing polyclonal and monoclonal antibodies are known to those of skill in the art (see, e.g., Paul (ed), *Fundamental Immunology, Third Edition* Raven Press, Ltd., NY (1993); Coligan, *Current Protocols in Immunology* Wiley/Greene, NY (1991); Harlow and Lane, *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY (1989); Stites *et al.* (eds.), *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY, (1986); and Kohler and Milstein, *Nature* 256:495-497 (1975)). Other suitable techniques for antibody preparation include selection of libraries of recombinant antibodies in phage or similar vectors (see, Huse *et al.*, *Science* 246:1275-1281 (1989); and Ward *et al.*, *Nature* 341:544-546 (1989)). Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 0.1 μM , preferably at least about 0.01 μM or better, and most typically and preferably, 0.001 μM or better.

The nucleic acids used in this invention can be either positive or negative probes. Positive probes bind to their targets and the presence of duplex formation is evidence of the presence of the target. Negative probes fail to bind to the suspect target and the absence of duplex formation is evidence of the presence of the target. For example, the use of a wild type specific nucleic acid probe or PCR primers may serve as a negative probe in an assay sample where only the nucleotide sequence of interest is present.

The sensitivity of the hybridization assays may be enhanced through use of a nucleic acid amplification system which multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBA[®], Cangene, Mississauga, Ontario) and Q Beta Replicase systems. These systems can be used to directly identify mutants where the PCR or LCR primers are designed to be extended or ligated only when a selected sequence is present. Alternatively, the selected sequences can be generally amplified using, for example, nonspecific PCR primers and the amplified target region later probed for a specific sequence indicative of a mutation.

A preferred embodiment is the use of allelic specific amplifications. In the case of PCR, the amplification primers are designed to bind to a portion of, for example, a gene encoding a G protein-coupled receptor protein, but the terminal base at the 3' end is used to discriminate between the mutant and wild-type forms of the G protein-coupled receptor gene. If the terminal base matches the point mutation or the wild-type, polymerase dependent three prime extension can proceed and an amplification product is detected. This method for detecting point mutations or polymorphisms is described in detail by Sommer *et al.*, in *Mayo Clin. Proc.* 64:1361-1372 (1989). By using appropriate controls, one can develop a kit having both positive and negative amplification products.

5 The products can be detected using specific probes or by simply detecting their presence or absence. A variation of the PCR method uses LCR where the point of discrimination, *i.e.*, either the point mutation or the wild-type bases fall between the LCR oligonucleotides. The ligation of the oligonucleotides becomes the means for discriminating between the mutant and wild-type forms of the gene encoding the G

10 protein-coupled receptor.

15

An alternative means for determining the level of expression of the nucleic acids of the present invention is *in situ* hybridization. *In situ* hybridization assays are well-known and are generally described in Angerer *et al.*, *Methods Enzymol.* 152:649-660 (1987). In an *in situ* hybridization assay, cells, preferentially human cells from the cerebellum or the hippocampus, are fixed to a solid support, typically a glass slide. If DNA is to be probed, the cells are denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of specific probes that are labeled. The probes are preferably labeled with radioisotopes or fluorescent reporters.

25 **VI. IMMUNOLOGICAL DETECTION OF THE GPCRs**

In numerous embodiments of the present invention, antibodies that specifically bind to the G protein-coupled receptors of the invention will be used. Such antibodies have numerous applications, including for the modulation of the activity of the G protein-coupled receptors and for immunoassays to detect the G protein-coupled receptors of the invention, as well as variants, derivatives, fragments, *etc.* thereof. Immunoassays can be used to qualitatively or quantitatively analyze the proteins of interest. A general overview of the applicable technology can be found in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Pubs., NY (1988).

Immunoassays for detecting target G protein-coupled receptor proteins are useful for diagnosing any disease or disorder in which GPCR-mediated signaling has been involved such as, e.g., Alzheimer's disease, depression, specific sarcomas and carcinomas, Parkinson's disease, psoriasis, rheumatoid arthritis, schizophrenia, tuberculosis, learning 5 and memory disorders, diabetes, reproduction and sex behavior disorders, anorexia, fat metabolism and body adiposity disorders, regulation of neurotransmitter release, pain perception, depression, regulation of hormone release, cardiovascular actions regulation, etc. In some embodiments, the antibodies of the present invention specifically bind to the G protein-coupled receptors of the invention and do not bind to other G protein-coupled 10 receptors or to G protein-coupled receptors from a different species, such as mouse, rat, etc. (identified GPCRs are listed in public databases, such as SwissProt, see <http://www.expasy.ch/sprot/sprot-top.html>, or GenBank, see <http://www.ncbi.nlm.nih.gov/>; see also *G protein coupled receptor Database*, <http://www.gcrdb.uthscsa.edu>). In some embodiments, the antibodies of the present 15 invention specifically bind to the galanin receptors of the invention and do not bind to other galanin receptors, such as GALR1, GALR2 and GALR3 (see, e.g., SwissProt accession numbers P47211, O43603, and O60755 for the sequences of the human GALR1, GALR2 and GALR3, respectively) or to galanin receptors from a different species (see, e.g., SwissProt accession numbers P56479, O88854, O88853, for the 20 sequences of the mouse GALR1, GALR2, and GALR3, respectively, and accession numbers Q62805, O08726, and O88626, for the sequences of the rat GALR1, GALR2, and GALR3, respectively).

A. Antibodies to Target Proteins

Methods for producing polyclonal and monoclonal antibodies that react 25 specifically with a protein of interest are known to those of skill in the art (see, e.g., Coligan, *supra*; and Harlow and Lane, *supra*; Stites *et al.*, *supra* and references cited therein; Goding, *supra*; and Kohler and Milstein, *Nature* 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors (see, Huse *et al.*, *supra*; and Ward *et 30 al.*, *supra*). For example, in order to produce antisera for use in an immunoassay, the protein of interest or an antigenic fragment thereof, is isolated as described herein. For example, a recombinant protein is produced in a transformed cell line. An inbred strain of mice or rabbits is immunized with the protein using a standard adjuvant, such as

Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used as an immunogen.

Polyclonal sera are collected and titered against the immunogen protein in
5 an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of 10^4 or greater are selected and tested for their cross-reactivity against non-G protein-coupled receptor proteins or even other homologous proteins from other organisms, using a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will
10 usually bind with a K_D of at least about 0.1 mM, more usually at least about 1 μM , preferably at least about 0.1 μM or better, and most preferably, 0.01 μM or better.

A number of proteins of the invention comprising immunogens may be used to produce antibodies specifically or selectively reactive with the proteins of interest. Recombinant protein is the preferred immunogen for the production of monoclonal or
15 polyclonal antibodies. Naturally occurring protein may also be used either in pure or impure form. Synthetic peptides made using the protein sequences described herein may also be used as an immunogen for the production of antibodies to the protein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells and purified as generally described *supra*. The product is then injected into an animal capable of
20 producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the protein.

Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. The animal's immune response to the immunogen
25 preparation is monitored by taking test bleeds and determining the titer of reactivity to the G protein-coupled receptor of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see*, Harlow and Lane, *supra*).

Monoclonal antibodies may be obtained using various techniques familiar to those of skill in the art. Typically, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*See*, Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include, *e.g.*, transformation with Epstein Barr Virus, oncogenes, or

retroviruses, or other methods well-known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a 5 vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *supra*.

Once target protein specific antibodies are available, the protein can be measured by a variety of immunoassay methods with qualitative and quantitative results 10 available to the clinician. For a review of immunological and immunoassay procedures in general, see, Stites, *supra*. Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Maggio, *Enzyme Immunoassay*, CRC Press, Boca Raton, Florida (1980); Tijssen, *supra*; and Harlow and Lane, *supra*.

15 Immunoassays to measure target proteins in a human sample may use a polyclonal antiserum which was raised to the protein partially encoded by a sequence described herein (e.g., a sequence selected from the sequences set forth in Table 1) or a fragment thereof. This antiserum is selected to have low cross-reactivity against non-G protein-coupled receptor proteins and any such cross-reactivity is removed by 20 immunoabsorption prior to use in the immunoassay.

Polyclonal antibodies that specifically bind to a G protein-coupled receptor of interest from a particular species can be made by subtracting out cross-reactive antibodies using G protein-coupled receptor homologs. In an analogous fashion, antibodies specific to a particular G protein-coupled receptor (e.g., a G protein-coupled 25 receptor encoded by a sequence set forth in Table 1) can be obtained in an organism with multiple G protein-coupled receptors genes by subtracting out cross-reactive antibodies using other G protein-coupled receptors.

Polyclonal antibodies that specifically bind to a galanin receptor of interest from a particular species can be made by subtracting out cross-reactive antibodies using 30 galanin receptor homologs. In an analogous fashion, antibodies specific to a particular galanin receptor (e.g., the galanin receptors of the invention) can be obtained in an organism with multiple galanin receptor genes by subtracting out cross-reactive antibodies using other galanin receptors, such as GALR1, GALR2 and GALR3.

B. Immunological Binding Assays

In a preferred embodiment, a protein of interest is detected and/or quantified using any of a number of well-known immunological binding assays (*see, e.g.*, U.S. Patent Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, *see also* Asai, *Methods in Cell Biology Volume 37: Antibodies in Cell Biology*, Academic Press, Inc. NY (1993); Stites, *supra*. Immunological binding assays (or immunoassays) typically utilize a “capture agent” to specifically bind to and often immobilize the analyte (in this case a G protein-coupled receptor of the invention or antigenic subsequences thereof). The capture agent is a moiety that specifically binds to the analyte. In a preferred embodiment, the capture agent is an antibody that specifically binds, for example, a GPCR of the invention. The antibody (*e.g.*, anti-GPCR antibody) may be produced by any of a number of means well-known to those of skill in the art and as described above.

Immunoassays also often utilize a labeling agent to specifically bind to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Thus, the labeling agent may be a labeled GPCR polypeptide or a labeled anti-GPCR antibody. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/protein complex.

In a preferred embodiment, the labeling agent is a second antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second antibody can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G, can also be used as the label agents. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (*see, generally*, Kronval *et al.* *J. Immunol.* 111:1401-1406 (1973); and Akerstrom *et al.*, *J. Immunol.* 135:2589-2542 (1985)).

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. The incubation time

will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

1. Non-competitive Assay Formats

5 Immunoassays for detecting proteins of interest from tissue samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured analyte (in this case the protein) is directly measured. In one preferred "sandwich" assay, for example, the capture agent (e.g., anti-GPCR antibodies) can be bound directly to a solid substrate where it is immobilized. These 10 immobilized antibodies then capture the G protein-coupled receptor present in the test sample. The G protein-coupled receptor thus immobilized is then bound by a labeling agent, such as a second anti-GPCR antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The 15 second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

2. Competitive Assay Formats

In competitive assays, the amount of target protein (analyte) present in the sample is measured indirectly by measuring the amount of an added (exogenous) analyte 20 (*i.e.*, a GPCR of interest) displaced (or competed away) from a capture agent (*i.e.*, anti-GPCR antibody) by the analyte present in the sample. In one competitive assay, a known amount of, in this case, the protein of interest is added to the sample and the sample is then contacted with a capture agent, in this case an antibody that specifically binds to the GPCR of interest. The amount of GPCR bound to the antibody is inversely proportional 25 to the concentration of GPCR present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of the GPCR bound to the antibody may be determined either by measuring the amount of subject protein present in a GPCR protein/antibody complex or, alternatively, by measuring the amount of remaining uncomplexed protein. The amount of GPCR protein may be 30 detected by providing a labeled GPCR protein molecule.

A hapten inhibition assay is another preferred competitive assay. In this assay, a known analyte, in this case the target protein, is immobilized on a solid substrate. A known amount of anti-GPCR antibody is added to the sample, and the sample is then contacted with the immobilized target. In this case, the amount of anti-GPCR antibody

bound to the immobilized GPCR is inversely proportional to the amount of GPCR protein present in the sample. Again, the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by 5 the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

Immunoassays in the competitive binding format can be used for cross-reactivity determinations. For example, the protein encoded by the sequences described herein can be immobilized on a solid support. Proteins are added to the assay which 10 compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to that of the protein encoded by any of the sequences described herein. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with each of the proteins listed above 15 are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the considered proteins, *e.g.*, distantly related homologs.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be 20 perhaps a protein of the present invention, to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required is less than 10 times the amount of the protein partially encoded by a sequence herein that is 25 required, then the second protein is said to specifically bind to an antibody generated to an immunogen consisting of the target protein.

3. Other Assay Formats

In a particularly preferred embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of a G protein-coupled receptor of the 30 invention in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support (such as, *e.g.*, a nitrocellulose filter, a nylon filter, or a derivatized nylon filter) and incubating the sample with the antibodies that specifically bind the protein of interest. For example, the anti-GPCR antibodies specifically bind to

the G protein-coupled receptor on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the antibodies against the protein of interest.

5 Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (*e.g.*, antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (*see*, Monroe *et al.*, *Amer. Clin. Prod. Rev.* 5:34-41 (1986)).

4. Reduction of Non-Specific Binding

10 One of skill in the art will appreciate that it is often desirable to use non-specific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of non-specific binding to the substrate. Means of reducing such non-specific binding are well-known to those of skill in the art. Typically, this involves coating the substrate with a
15 proteinaceous composition. In particular, protein compositions, such as bovine serum albumin (BSA), nonfat powdered milk and gelatin, are widely used.

5. Labels

20 The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific
25 binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most labels useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical,
30 optical or chemical means. Useful labels in the present invention include magnetic beads (*e.g.*, Dynabeads™), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (*e.g.*, ^3H , ^{125}I , ^{35}S , ^{14}C , or ^{32}P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, *etc.*) beads.

The label may be coupled directly or indirectly to the desired component of the assay according to methods well-known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on the sensitivity

required, the ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

Non-radioactive labels are often attached by indirect means. The molecules can also be conjugated directly to signal generating compounds, e.g., by 5 conjugation with an enzyme or fluorescent compound. A variety of enzymes and fluorescent compounds can be used with the methods of the present invention and are well-known to those of skill in the art (for a review of various labeling or signal producing systems which may be used, see, e.g., U.S. Patent No. 4,391,904).

Means of detecting labels are well-known to those of skill in the art. Thus, 10 for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such 15 as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple colorimetric labels may be detected directly by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads 20 appear the color of the bead.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target 25 antibodies. In this format, none of the components need to be labeled and the presence of the target antibody is detected by simple visual inspection.

VII. SCREENING FOR MODULATORS OF THE GPCRs OF THE INVENTION

The invention also provides methods for identifying compounds that modulate signaling mediated by the G protein-coupled receptors of the invention. These 30 compounds include both those that modulate the expression and those that modulate the activity of the G protein-coupled receptors of the invention. Furthermore, these compounds may modulate the expression and/or activity of one or of various G protein-coupled receptors of the invention, and optionally of all the G protein-coupled receptors

of the invention. In addition, the identified compounds can also modulate, *e.g.*, the development of Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, sarcomas such as, chondrosarcoma, Ewing's 5 sarcoma, and osteosarcoma, carcinomas such as, basal cell carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, and thyroid carcinoma, psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma 10 multiform, Hodgkin's disease, lymphoma, macular degeneration, malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, learning and memory processes, reproduction and sex behavior, feeding behavior, fat metabolism and body adiposity, neurotransmitter release, pain perception, depression, hormone release, cardiovascular actions, or any other disease or disorder 15 involving GPCR-mediated signaling.

A. Screening for Modulators of the G Protein-Coupled Receptors

The present invention provides methods for identifying compounds that increase or decrease the expression level or the activity of one or more G protein-coupled receptors of interest. Compounds that are identified as modulators of the expression or 20 activity of one or more G protein-coupled receptors of the invention using the methods described herein find use both *in vitro* and *in vivo*. For example, one can treat cell cultures with the modulators in experiments designed to determine the mechanisms by which GPCR-mediated signaling is regulated. Compounds that modulate the activity of the G protein-coupled receptors are useful for studying, for example, the mechanisms that 25 lead to depression, Alzheimer's disease, specific sarcomas and carcinomas, other cancers such as lymphomas and melanomas, psoriasis, cardiomyopathies, *etc.* Compounds that modulate the activity of the galanin receptor are useful for studying, for example, the mechanisms that lead to growth hormone release, depression or fat accumulation, neurotransmitter or insulin release.

30 The methods for isolating compounds that modulate the expression of the G protein-coupled receptors of the invention typically involve culturing a cell in the presence of a potential modulator to form a first cell culture. RNA (or cDNA) from the first cell culture is contacted with one or more probes, each probe comprising a

polynucleotide sequence encoding a G protein-coupled receptor of the invention (e.g., a nucleotide sequence selected from the group of sequences set forth in Table 1). The amount of the probe(s) which hybridizes to the RNA (or cDNA) from the first cell culture is determined. Typically, one determines whether the amount of the probe(s) which

5 hybridizes to the RNA (or cDNA) is increased or decreased relative to the amount of the probe(s) which hybridizes to RNA (or cDNA) from a second cell culture grown in the absence of the modulator.

The G protein-coupled receptors of the invention and their alleles and polymorphic variants mediate signaling in different pathways involving a variety of ligands. The activity of G protein-coupled receptor polypeptides can be assessed using a variety of *in vitro* and *in vivo* assays to determine functional, chemical, and physical effects, e.g., measuring ligand binding (e.g., radioactive ligand binding), second messengers (e.g., cAMP, cGMP, IP₃, DAG, or Ca²⁺), ion flux, phosphorylation levels, transcription levels, neurotransmitter levels, and the like. Furthermore, such assays can be used to test for inhibitors and activators of the G protein-coupled receptors of the invention. Modulators can also be genetically altered versions of the present G protein-coupled receptors. Such modulators of GPCR-mediated signaling activity are useful for treating a variety of diseases and disorders described herein. For a general review of GPCR signal transduction and methods of assaying signal transduction, see, e.g., *Methods in Enzymology* vols. 237 and 238 (1994) and volume 96 (1983); Bourne *et al.*, *Nature* 10:349:117-27 (1991); Bourne *et al.*, *Nature* 348:125-32 (1990); Pitcher *et al.*, *Annu. Rev. Biochem.* 67:653-92 (1998).

The G protein-coupled receptors of the assay will typically be polypeptides having identity with polypeptides encoded by a nucleic acid molecule having a nucleotide sequence selected from the sequences set forth in Table 1, or conservatively modified variants thereof.

Generally, the amino acid sequence identity will be at least 70%, 75%, 80%, 85%, 90%, 95% or more identity and further will not be identical to the sequences for known GPCRs (for sequences of identified GPCRs, see, e.g.,

30 <http://www.gcrdb.uthscsa.edu>; <http://www.ncbi.nlm.nih.gov>; and <http://www.expasy.ch/sprot/sprot.top.html>). With regard to galanin receptors, the amino acid sequences of the invention will not be identical to the sequences for GALR1, GALR2 or GALR3 (see, e.g., SwissProt accession numbers P47211, O43603, and O60755 for the sequences of the human GALR1, GALR2 and GALR3, respectively).

Optionally, the polypeptide(s) of the assays will comprise a domain of a G protein-coupled receptor, such as an extracellular domain, transmembrane region, transmembrane domain, cytoplasmic domain, ligand binding domain, subunit association domain, active site, and the like. The polypeptides of the present invention may also be 5 polypeptides comprising a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85%, and most preferably 90% or more, identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1, and having substantially 10 the same biological activity. Either the G protein-coupled receptor protein or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.

Modulators of the activity of G protein-coupled receptors are tested using G protein-coupled receptors polypeptides as described above, either recombinant or 15 naturally occurring. The proteins can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal, either recombinant or naturally occurring. For example, neurons, transformed cells, or membranes can be used. Modulation is tested using one of the *in vitro* or *in vivo* assays described herein. G protein-mediated signaling can also be examined *in vitro* with soluble or solid state 20 reactions, using a full-length G protein-coupled receptor or a chimeric molecule such as an extracellular domain or transmembrane region, or combination thereof, of a G protein-coupled receptor covalently linked to a heterologous signal transduction domain, or a heterologous extracellular domain and/or transmembrane region covalently linked to the transmembrane and/or cytoplasmic domain of a G protein-coupled receptor. 25 Furthermore, ligand-binding domains of the protein of interest can be used *in vitro* in soluble or solid state reactions to assay for ligand binding. In numerous embodiments, a chimeric receptor will be made that comprises all or part of a G protein-coupled receptor polypeptide as well as an additional sequence that facilitates the localization of the G protein-coupled receptor to the membrane.

30 Ligand binding to a G protein-coupled receptor, a domain thereof, or a chimeric protein can be tested in solution, in a bilayer membrane, attached to a solid phase, in a lipid monolayer, or in vesicles. Binding of a modulator can be tested using, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index) hydrodynamic (e.g., shape), chromatographic, or solubility properties.

G protein-coupled receptor-G protein interactions can also be examined. For example, binding of the G protein to the receptor or its release from the receptor can be examined. For example, in the absence of GTP, an activator will lead to the formation of a tight complex of a G protein (all three subunits) with the receptor. This complex can 5 be detected in a variety of ways. Such an assay can be modified to search for inhibitors, e.g., by adding an activator to the G protein-coupled receptor and G protein in the absence of GTP, which form a tight complex, and then screen for inhibitors by looking at dissociation of the G protein-coupled receptor-G protein complex. In the presence of GTP, release of the alpha subunit of the G protein from the other two G protein subunits 10 serves as a criterion of activation.

In some embodiments, G protein-coupled receptors-ligand interactions are monitored as a function of G protein-coupled receptors activation.

An activated or inhibited G protein will in turn alter the properties of target enzymes, channels, and other effector proteins. Target enzymes and effector proteins for 15 G protein-coupled receptors that can be used in the context of the present invention are known to those of skill in the art.

In some embodiments, a G protein-coupled receptor polypeptide is expressed in a eukaryotic cell as a chimeric receptor with a heterologous, chaperone sequence that facilitates its maturation and targeting through the secretory pathway. 20 Chimeric G protein-coupled receptors can be expressed in any eukaryotic cell, such as HEK-293 cells. Preferably, the cells comprise a functional G protein that is capable of coupling the chimeric receptor to an intracellular signaling pathway or to a signaling protein. Activation of such chimeric receptors in such cells can be detected using any standard method, such as by detecting changes in intracellular calcium by detecting 25 FURA-2 dependent fluorescence in the cell.

In addition, activated G protein-coupled receptors become substrates for kinases. Phosphorylation of the G protein-coupled receptors of the invention can thus also be measured as a means to detect activation of the receptors. Phosphorylation may be detected by assaying the transfer of ³²P from gamma-labeled GTP to the receptor with 30 a scintillation counter.

Samples or assays that are treated with a potential G protein-coupled receptor inhibitor or activator are compared to control samples without the test compound, to examine the extent of modulation. Such assays may be carried out in the presence of ligand, and modulation of the ligand-dependent activation is monitored.

Control samples (untreated with activators or inhibitors) are assigned a relative G protein-coupled receptor activity value of 100. Inhibition of a G protein-coupled receptor protein is achieved when the G protein-coupled receptor activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of a G protein-coupled receptor protein is achieved when the G protein-coupled receptor activity value relative to the control is 110%, optionally 150%, 200-500%, or 1000-2000% or more.

Changes in ion flux may be assessed by determining changes in polarization (*i.e.*, electrical potential) of the cell or membrane expressing a G protein-coupled receptor of interest. One means to determine changes in cellular polarization is by measuring changes in current (thereby measuring changes in polarization) with voltage-clamp and patch-clamp techniques, *e.g.*, the "cell-attached" mode, the "inside-out" mode, and the "whole cell" mode (*see, e.g.*, Ackerman *et al.*, *New Engl. J. Med.* 336:1575-1595 (1997)). Whole cell currents are conveniently determined using the standard methodology (*see, e.g.*, Hamil *et al.*, *Pflugers Archiv.* 391:85 (1981)). Other known assays include: radiolabeled ion flux assays and fluorescence assays using voltage-sensitive dyes (*see, e.g.*, Vestergaard-Bogind *et al.*, *J. Membrane Biol.* 88:67-75 (1988); Gonzales & Tsien, *Chem. Biol.* 4:269-277 (1997); Daniel *et al.*, *J. Pharmacol. Meth.* 25:185-193 (1991); Holevinsky *et al.*, *J. Membrane Biology* 137:59-70 (1994)). Generally, the compounds to be tested are present in the range from 1 pM to 100 mM.

The effects of the test compounds upon the function of the polypeptides can be measured by examining any of the parameters described above, and other parameters known to those of skill in the art. Any suitable physiological change that affects G protein-coupled receptor activity can be used to assess the influence of a test compound on the G protein-coupled receptors of this invention. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as transmitter release, hormone release, transcriptional changes to both known and uncharacterized genetic markers, changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as Ca²⁺, IP3, cGMP, or cAMP.

Preferred assays for G protein-coupled receptors include cells that are loaded with ion or voltage sensitive dyes to report receptor activity. Assays for determining activity of such receptors can also use known agonists and antagonists for other G protein-coupled receptors as negative or positive controls to assess activity of tested compounds. In assays for identifying modulatory compounds (*e.g.*, agonists,

antagonists), changes in the level of ions in the cytoplasm or membrane voltage will be monitored using an ion sensitive or membrane voltage fluorescent indicator, respectively. Among the ion-sensitive indicators and voltage probes that may be employed are those disclosed in the Molecular Probes 1997 Catalog. For G protein-coupled receptors, 5 promiscuous G proteins can be used in the assay of choice (Wilkie *et al.*, *Proc. Natl. Acad. Sci. USA* 88:10049-10053 (1991)). Such promiscuous G proteins allow coupling of a wide range of receptors.

Other assays to determine the activity of G protein-coupled receptors, can involve measuring changes in the level of intracellular cyclic nucleotides, *e.g.*, cAMP or 10 cGMP, that occur due to the activation or inhibition of enzymes such as adenylyl cyclase upon activation of the receptor.

In one embodiment, the changes in intracellular cAMP or cGMP can be measured using immunoassays. The method described in Offermanns & Simon, *J. Biol. Chem.* 270:15175-15180 (1995) may be used to determine the level of cAMP. Also, the 15 method described in Felley-Bosco *et al.*, *Am. J. Resp. Cell and Mol. Biol.* 11:159-164 (1994) may be used to determine the level of cGMP. Further, an assay kit for measuring cAMP and/or cGMP is described in U.S. Patent No. 4,115,538.

In another embodiment, transcription levels can be measured to assess the effects of a test compound on signal transduction. A host cell containing a G protein- 20 coupled receptor of interest is contacted with a test compound for a sufficient time to effect any interactions, and then the level of gene expression is measured. The amount of time to effect such interactions may be empirically determined, such as by running a time course and measuring the level of transcription as a function of time. The amount of transcription may be measured by using any method known to those of skill in the art to 25 be suitable. For example, mRNA expression of the protein of interest may be detected using northern blots or their polypeptide products may be identified using immunoassays. Alternatively, transcription based assays using reporter gene may be used as described in U.S. Patent No. 5,436,128. The reporter genes can be, *e.g.*, chloramphenicol acetyltransferase, luciferase, β -galactosidase and alkaline phosphatase. Furthermore, the 30 protein of interest can be used as an indirect reporter via attachment to a second reporter such as green fluorescent protein (*see, e.g.*, Mistili and Spector, *Nature Biotechnology* 15:961-964 (1997)). The amount of transcription is then compared to the amount of transcription in either the same cell in the absence of the test compound, or it may be

compared with the amount of transcription in a substantially identical cell that lacks the protein of interest. A substantially identical cell may be derived from the same cells from which the recombinant cell was prepared but which had not been modified by introduction of heterologous DNA. Any difference in the amount of transcription 5 indicates that the test compound has in some manner altered the activity of the protein of interest.

Any other method that allows to determine the effect of a compounds on the activity of a G protein-coupled receptor of interest can also be used in the context of the present invention (for articles disclosing methods for determining the activity of G 10 protein-coupled receptors, see, e.g., Fisone *et al.*, *Brain Res.* 568:279-84 (1991); Ogren *et al.*, *Ann. NY Acad. Sci.* 863:342-63 (1998); Wang *et al.*, *Neuropeptides* 33:197-205 (1999)).

B. Modulators of the Activity of the G Protein-Coupled Receptors of the Invention

15 The compounds tested as modulators of the G protein-coupled receptors of the invention can be any small chemical compound, or a biological entity, such as a protein, sugar, nucleic acid or lipid. Alternatively, modulators can be genetically altered versions of a G protein-coupled receptor gene. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a 20 potential modulator or ligand in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates 25 in robotic assays). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

In one preferred embodiment, high throughput screening methods involve 30 providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator or ligand compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus

identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining 5 a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*, the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building 10 blocks.

Preparation and screening of combinatorial chemical libraries is well-known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.*, U.S. Patent No. 5,010,175; Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991); and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other 15 chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to, peptoids (*e.g.*, PCT Publication No. WO 91/19735), encoded peptides (*e.g.*, PCT Publication WO 93/20242), random bio-oligomers (*e.g.*, PCT Publication No. WO 92/00091), benzodiazepines (*e.g.*, U.S. Patent No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding 20 (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl 25 phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see* Ausubel *et al.*, Berger *et al.*, and Sambrook *et al.*, all *supra*), peptide nucleic acid libraries (*see, e.g.*, U.S. Patent No. 5,539,083), antibody libraries (*see, e.g.*, Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.*, Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent No. 5,593,853), small 30 organic molecule libraries (*see, e.g.*, benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent No. 5,569,588; thiazolidinones and metathiazanones, U.S. Patent No. 5,549,974; pyrrolidines, U.S. Patent Nos. 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent No. 5,506,337; benzodiazepines, 5,288,514, and the like), *etc.*.

Devices for the preparation of combinatorial libraries are commercially available (see, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves 5 commercially available (see, e.g., ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, MO, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, etc.).

C. Solid State and Soluble High Throughput Assays

In one embodiment, the invention provides soluble assays using molecules such as a domain, such as a ligand binding domain, an extracellular domain, a 10 transmembrane domain (e.g., one comprising seven transmembrane regions and cytosolic loops), the transmembrane domain and a cytoplasmic domain, an active site, a subunit association region, etc., a domain that is covalently linked to a heterologous protein to create a chimeric molecule, a G protein-coupled receptor, or a cell or tissue expressing a G protein-coupled receptor, either naturally occurring or recombinant. In another 15 embodiment, the invention provides solid phase based *in vitro* assays in a high throughput format, where the domain, chimeric molecule, G protein-coupled receptor, or cell or tissue expressing the G protein-coupled receptor is attached to a solid phase substrate.

In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well 20 of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (e.g., 96) modulators. If 1536 well plates are used, then a single plate can easily 25 assay from about 100 to about 1500 different compounds. It is possible to assay several different plates per day. Assay screens for up to about 6,000-20,000 different compounds are possible using the integrated systems of the invention. More recently, microfluidic approaches to reagent manipulation have been developed.

The molecule of interest can be bound to the solid state component, directly or indirectly, via covalent or non covalent linkage, e.g., via a tag. The tag can be 30 any of a variety of components. In general, a molecule which binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest (e.g., the G protein-coupled receptor of interest) is attached to the solid support by interaction of the tag and the tag binder.

A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, etc.) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders (see, SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody which recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs, such as agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherin family, the integrin family, the selectin family, and the like; see, e.g., Pigott and Power, *The Adhesion Molecule Facts Book I* (1993)). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), intracellular receptors (e.g., which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to those of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulphydryl linkages, or heterofunctional linkages.

Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent which fixes a chemical group to the surface which is reactive with a portion of the tag binder. For example,

5 groups which are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature (*see, e.g.*, Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, *e.g.*,

10 peptides); Geysen *et al.*, *J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank and Doring, *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al.*, *Science* 251:767-777 (1991); Sheldon *et al.*, *Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al.*, *Nature Medicine* 2(7):753759 (1996) (all describing arrays of

15 biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

The invention provides *in vitro* assays for identifying, in a high throughput format, compounds that can modulate the expression or activity of the G protein-coupled receptors of the invention. Control reactions that measure the G protein-coupled receptor activity of the cell in a reaction that does not include a potential modulator are optional, as the assays are highly uniform. Such optional control reactions are appropriate and increase the reliability of the assay. Accordingly, in a preferred embodiment, the methods of the invention include such a control reaction. For each of the assay formats described,

20 "no modulator" control reactions which do not include a modulator provide a background level of binding activity.

In some assays it will be desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. First, a known activator of the G protein-coupled receptors of the invention can be incubated with one sample of the assay, and the resulting increase in signal resulting from an increased expression level or activity of a G protein-coupled receptor determined according to the methods herein. Second, a known inhibitor of the G protein-coupled receptors of the invention can be added, and the resulting decrease in signal for the expression or activity of a G protein-coupled receptor similarly detected. It

will be appreciated that modulators can also be combined with activators or inhibitors to find modulators which inhibit the increase or decrease that is otherwise caused by the presence of the known modulator of the G protein-coupled receptor.

D. Computer-Based Assays

5 Yet another assay for compounds that modulate the activity of G protein-coupled receptors involves computer assisted drug design, in which a computer system is used to generate a three-dimensional structure of a G protein-coupled receptor based on the structural information encoded by its amino acid sequence. The input amino acid sequence interacts directly and actively with a pre-established algorithm in a computer 10 program to yield secondary, tertiary, and quaternary structural models of the protein. The models of the protein structure are then examined to identify regions of the structure that have the ability to bind, *e.g.*, ligands. These regions are then used to identify ligands that bind to the protein.

The three-dimensional structural model of the protein is generated by 15 entering protein amino acid sequences of at least 10 amino acid residues (or corresponding nucleic acid sequences encoding a G protein-coupled receptor) into the computer system. The nucleotide sequence encoding the GPCR can be any sequence encoding a polypeptide having at least 30%, optionally at least 40%, 50%, 60%, 70%, 80%, 90% or more identity with a polypeptide encoded by a nucleic acid molecule having 20 a sequence selected from the group consisting of the sequences set forth in Table 1, and conservatively modified versions thereof. The amino acid sequences encoded by the nucleic acid sequences provided herein represent the primary sequences or subsequences of the proteins, which encode the structural information of the proteins. At least 10 residues of an amino acid sequence (or a nucleotide sequence encoding 10 amino acids) 25 are entered into the computer system from computer keyboards, computer readable substrates that include, but are not limited to, electronic storage media (*e.g.*, magnetic diskettes, tapes, cartridges, and chips), optical media (*e.g.*, CD ROM), information distributed by internet sites, and by RAM. The three-dimensional structural model of the protein is then generated by the interaction of the amino acid sequence and the computer 30 system, using software known to those of skill in the art.

The amino acid sequence represents a primary structure that encodes the information necessary to form the secondary, tertiary and quaternary structures of the protein of interest. The software looks at certain parameters encoded by the primary

sequence to generate the structural model. These parameters are referred to as "energy terms" and primarily include electrostatic potentials, hydrophobic potentials, solvent accessible surfaces, and hydrogen bonding. Secondary energy terms include van der Waals potentials. Biological molecules form the structures that minimize the energy
5 terms in a cumulative fashion. The computer program uses these terms encoded by the primary structure or amino acid sequence to create the secondary structural model.

The tertiary structure of the protein encoded by the secondary structure is then formed on the basis of the energy terms of the secondary structure. The user at this point can enter additional variables such as whether the protein is membrane bound or
10 soluble, its location in the body, and its cellular location, *e.g.*, cytoplasmic, surface, or nuclear. These variables along with the energy terms of the secondary structure are used to form the model of the tertiary structure. In modeling the tertiary structure, the computer program matches hydrophobic faces of secondary structure with like, and hydrophilic faces of secondary structure with like.

15 Once the structure has been generated, potential ligand-binding regions are identified by the computer system. Three-dimensional structures for potential ligands are generated by entering amino acid or nucleotide sequences or chemical formulas of compounds, as described above. The three-dimensional structure of the potential ligand is then compared to that of the G protein-coupled receptor to identify ligands that bind to
20 the protein. Binding affinity between the protein and ligands is determined using energy terms to determine which ligands have an enhanced probability of binding to the protein.

Computer systems are also used to screen for mutations, polymorphic variants, alleles and interspecies homologs of genes encoding the G protein-coupled receptors of the invention. Such mutations can be associated with disease states or
25 genetic traits. As described above, GeneChip™ and related technology can also be used to screen for mutations, polymorphic variants, alleles and interspecies homologs. Once the variants are identified, diagnostic assays can be used to identify patients having such mutated genes. Identification of the mutated G protein-coupled receptor genes involves receiving input of a first amino acid sequence of a G protein-coupled receptor (or of a
30 first nucleic acid sequence encoding a GPCR of the invention), *e.g.*, any amino acid sequence having at least 30%, optionally at least 40%, 50%, 60%, 70%, 80%, 90% or more identity with a polypeptide encoded by a nucleic acid molecule having a sequence selected from the group consisting of the sequences set forth in Table 1, or conservatively

modified versions thereof, or alternatively any amino acid sequence comprising a region of 15 amino acids or more, optionally 30 amino acids or more, having at least 80%, preferably at least 85%, and most preferably 90% or more, identity with a region of 15 amino acids or more, optionally 30 amino acids or more, from a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence selected from the group consisting of the sequences set forth in Table 1. The sequence is entered into the computer system as described above. The first nucleic acid or amino acid sequence is then compared to a second nucleic acid or amino acid sequence that has substantial identity to the first sequence. The second sequence is entered into the computer system in the manner described above. Once the first and second sequences are compared, nucleotide or amino acid differences between the sequences are identified. Such sequences can represent allelic differences in various G protein-coupled receptor genes, and mutations associated with disease states and genetic traits.

VIII. COMPOSITIONS, KITS AND INTEGRATED SYSTEMS

The invention provides compositions, kits and integrated systems for practicing the assays described herein using nucleic acids encoding the G protein-coupled receptors of the invention, or the G protein-coupled receptors proteins themselves, anti-G protein-coupled receptors antibodies, etc.

The invention provides assay compositions for use in solid phase assays; such compositions can include, for example, one or more nucleic acids encoding a G protein-coupled receptor immobilized on a solid support, and a labeling reagent. In each case, the assay compositions can also include additional reagents that are desirable for hybridization. Modulators of expression or activity of a G protein-coupled receptor of the invention can also be included in the assay compositions.

The invention also provides kits for carrying out the assays of the invention. The kits typically include a probe that comprises a polynucleotide sequence encoding a G protein-coupled receptor, and a label for detecting the presence of the probe. The kits may include several polynucleotide sequences encoding G protein-coupled receptors of the invention. Kits can include any of the compositions noted above, and optionally further include additional components such as instructions to practice a high-throughput method of assaying for an effect on expression of the genes encoding the G protein-coupled receptors of the invention, or on activity of the G protein-coupled receptors of the invention, one or more containers or compartments (e.g., to hold the

probe, labels, or the like), a control modulator of the expression or activity of G protein-coupled receptors, a robotic armature for mixing kit components or the like.

The invention also provides integrated systems for high-throughput screening of potential modulators for an effect on the expression or activity of the G 5 protein-coupled receptors of the invention. The systems typically include a robotic armature which transfers fluid from a source to a destination, a controller which controls the robotic armature, a label detector, a data storage unit which records label detection, and an assay component such as a microtiter dish comprising a well having a reaction mixture or a substrate comprising a fixed nucleic acid or immobilization moiety.

10 A number of robotic fluid transfer systems are available, or can easily be made from existing components. For example, a Zymate XP (Zymark Corporation; Hopkinton, MA) automated robot using a Microlab 2200 (Hamilton; Reno, NV) pipetting station can be used to transfer parallel samples to 96 well microtiter plates to set up several parallel simultaneous STAT binding assays.

15 Optical images viewed (and, optionally, recorded) by a camera or other recording device (e.g., a photodiode and data storage device) are optionally further processed in any of the embodiments herein, e.g., by digitizing the image and storing and analyzing the image on a computer. A variety of commercially available peripheral equipment and software is available for digitizing, storing and analyzing a digitized video 20 or digitized optical image, e.g., using PC (Intel x86 or Pentium chip-compatible DOS®, OS2® WINDOWS®, WINDOWS NT®, WINDOWS95® or WINDOWS98® based computers), MACINTOSH®, or UNIX® based (e.g., SUN® work station) computers.

One conventional system carries light from the specimen field to a cooled 25 charge-coupled device (CCD) camera, in common use in the art. A CCD camera includes an array of picture elements (pixels). The light from the specimen is imaged on the CCD. Particular pixels corresponding to regions of the specimen (e.g., individual hybridization sites on an array of biological polymers) are sampled to obtain light intensity readings for each position. Multiple pixels are processed in parallel to increase speed. The apparatus and methods of the invention are easily used for viewing any sample, e.g., by fluorescent 30 or dark field microscopic techniques.

IX. GENE THERAPY APPLICATIONS

A variety of human diseases can be treated by therapeutic approaches that involve stably introducing a gene into a human cell such that the gene is transcribed and

the gene product is produced in the cell. Diseases amenable to treatment by this approach include inherited diseases, including those in which the defect is in a single gene. Gene therapy is also useful for treatment of acquired diseases and other conditions. For discussions on the application of gene therapy towards the treatment of genetic as well as 5 acquired diseases, see, Miller, *Nature* 357:455-460 (1992); and Mulligan, *Science* 260:926-932 (1993).

In the context of the present invention, gene therapy can be used for treating a variety of disorders and/or diseases in which G protein-coupled receptor-mediated signaling has been implicated. For example, introduction by gene therapy of 10 polynucleotides encoding a G protein-coupled receptor of the invention can be used to treat, e.g., Alzheimer's disease, rheumatoid arthritis, osteoarthritis, osteoporosis, amyotrophic lateral sclerosis, multiple sclerosis and atherosclerosis, asthma, depression, epilepsy, schizophrenia, Parkinson's disease, a number of sarcomas (e.g., chondrosarcoma, Ewing's sarcoma, osteosarcoma, etc.) and carcinomas (e.g., basal cell 15 carcinoma, breast carcinoma, embryonal carcinoma, ovarian carcinoma, renal cell carcinoma, lung adenocarcinoma, lung small cell carcinoma, pancreatic carcinoma, prostate carcinoma, transitional carcinoma of the bladder, squamous cell carcinoma, thyroid carcinoma, etc.), psoriasis, cardiomyopathy, Crohn's disease, Duchenne muscular dystrophy, glioblastoma multiform, Hodgkin's disease, lymphoma, macular degeneration, 20 malignant fibrous histiocytoma, melanoma, meningioma, mesothelioma, seminoma, tuberculosis, tonsil, ulcerative colitis, etc. Introduction by gene therapy of polynucleotides encoding a galanin receptor of the invention can be used to treat, e.g., anorexia, to induce nerve regeneration and to decrease noninception. In addition, antisense 25 polynucleotides can also be administered using gene therapy to treat, e.g., obesity, diabetes

A. Vectors for Gene Delivery

For delivery to a cell or organism, the nucleic acids of the invention can be incorporated into a vector. Examples of vectors used for such purposes include expression plasmids capable of directing the expression of the nucleic acids in the target 30 cell. In other instances, the vector is a viral vector system wherein the nucleic acids are incorporated into a viral genome that is capable of transfecting the target cell. In a preferred embodiment, the nucleic acids can be operably linked to expression and control

sequences that can direct expression of the gene in the desired target host cells. Thus, one can achieve expression of the nucleic acid under appropriate conditions in the target cell.

B. Gene Delivery Systems

Viral vector systems useful in the expression of the nucleic acids include, 5 for example, naturally occurring or recombinant viral vector systems. Depending upon the particular application, suitable viral vectors include replication competent, replication deficient, and conditionally replicating viral vectors. For example, viral vectors can be derived from the genome of human or bovine adenoviruses, vaccinia virus, herpes virus, adeno-associated virus, minute virus of mice (MVM), HIV, sindbis virus, and retroviruses 10 (including, but not limited to, Rous sarcoma virus), and MoMLV. Typically, the genes of interest are inserted into such vectors to allow packaging of the gene construct, typically with accompanying viral DNA, followed by infection of a sensitive host cell and expression of the gene of interest.

As used herein, "gene delivery system" refers to any means for the 15 delivery of a nucleic acid of the invention to a target cell. In some embodiments of the invention, nucleic acids are conjugated to a cell receptor ligand for facilitated uptake (e.g., invagination of coated pits and internalization of the endosome) through an appropriate linking moiety, such as a DNA linking moiety (see, e.g., Wu *et al.*, *J. Biol. Chem.* 263:14621-14624 (1988); and WO 92/06180). For example, nucleic acids can be 20 linked through a polylysine moiety to asialo-oromucocid, which is a ligand for the asialoglycoprotein receptor of hepatocytes.

Similarly, viral envelopes used for packaging gene constructs that include 25 the nucleic acids of the invention can be modified by the addition of receptor ligands or antibodies specific for a receptor to permit receptor-mediated endocytosis into specific cells (see, e.g., WO 93/20221; WO 93/14188; and WO 94/06923). In some embodiments of the invention, the DNA constructs of the invention are linked to viral proteins, such as adenovirus particles, to facilitate endocytosis (Curiel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88:8850-8854 (1991)). In other embodiments, molecular conjugates of the instant 30 invention can include microtubule inhibitors (WO 94/06922), synthetic peptides mimicking influenza virus hemagglutinin (Plank *et al.*, *J. Biol. Chem.* 269:12918-12924 (1994)), and nuclear localization signals such as SV40 T antigen (WO 93/19768).

Retroviral vectors are also useful for introducing the nucleic acids of the invention into target cells or organisms. Retroviral vectors are produced by genetically

manipulating retroviruses. The viral genome of retroviruses is RNA. Upon infection, this genomic RNA is reverse transcribed into a DNA copy which is integrated into the chromosomal DNA of transduced cells with a high degree of stability and efficiency. The integrated DNA copy is referred to as a provirus and is inherited by daughter cells as is any other gene. The wild type retroviral genome and the proviral DNA have three genes, the *gag*, the *pol* and the *env* genes, which are flanked by two long terminal repeat (LTR) sequences. The *gag* gene encodes the internal structural (nucleocapsid) proteins; the *pol* gene encodes the RNA directed DNA polymerase (reverse transcriptase); and the *env* gene encodes viral envelope glycoproteins. The 5' and 3' LTRs serve to promote transcription and polyadenylation of virion RNAs. Adjacent to the 5' LTR are sequences necessary for reverse transcription of the genome (the tRNA primer binding site) and for efficient encapsulation of viral RNA into particles (the Psi site) (*see*, Mulligan, *In: Experimental Manipulation of Gene Expression*, Inouye (ed), 155-173 (1983); Mann *et al.*, *Cell* 33:153-159 (1983); Cone and Mulligan, *Proc. Natl. Acad. Sci. U.S.A.* 81:6349-6353 (1984)).

The design of retroviral vectors is well-known to those of ordinary skill in the art. In brief, if the sequences necessary for encapsidation (or packaging of retroviral RNA into infectious virions) are missing from the viral genome, the result is a *cis* acting defect which prevents encapsidation of genomic RNA. However, the resulting mutant is still capable of directing the synthesis of all virion proteins. Retroviral genomes from which these sequences have been deleted, as well as cell lines containing the mutant genome stably integrated into the chromosome are well-known in the art and are used to construct retroviral vectors. Preparation of retroviral vectors and their uses are described in many publications including, *e.g.*, European Patent Application EPA 0 178 220; U.S. Patent No. 4,405,712; Gilboa, *Biotechniques* 4:504-512 (1986); Mann *et al.*, *Cell* 33:153-159 (1983); Cone and Mulligan, *Proc. Natl. Acad. Sci. USA* 81:6349-6353 (1984); Eglitis *et al.*, *Biotechniques* 6:608-614 (1988); Miller *et al.*, *Biotechniques* 7:981-990 (1989); Miller (1992) *supra*; Mulligan (1993), *supra*; and WO 92/07943.

The retroviral vector particles are prepared by recombinantly inserting the desired nucleotide sequence into a retrovirus vector and packaging the vector with retroviral capsid proteins by use of a packaging cell line. The resultant retroviral vector particle is incapable of replication in the host cell but is capable of integrating into the host cell genome as a proviral sequence containing the desired nucleotide sequence. As a

result, the patient is capable of producing, for example, a G protein-coupled receptor of interest and thus restore the cells to a normal phenotype.

Packaging cell lines that are used to prepare the retroviral vector particles are typically recombinant mammalian tissue culture cell lines that produce the necessary viral structural proteins required for packaging, but which are incapable of producing infectious virions. The defective retroviral vectors that are used, on the other hand, lack these structural genes but encode the remaining proteins necessary for packaging. To prepare a packaging cell line, one can construct an infectious clone of a desired retrovirus in which the packaging site has been deleted. Cells comprising this construct will express all structural viral proteins, but the introduced DNA will be incapable of being packaged. Alternatively, packaging cell lines can be produced by transforming a cell line with one or more expression plasmids encoding the appropriate core and envelope proteins. In these cells, the *gag*, *pol*, and *env* genes can be derived from the same or different retroviruses.

A number of packaging cell lines suitable for the present invention are also available in the prior art. Examples of these cell lines include Crip, GPE86, PA317 and PG13 (*see* Miller *et al.*, *J. Virol.* 65:2220-2224 (1991)). Examples of other packaging cell lines are described in Cone and Mulligan, *Proc. Natl. Acad. Sci. USA* 81:6349-6353 (1984); Danos and Mulligan, *Proc. Natl. Acad. Sci. USA* 85:6460-6464 (1988); Eglitis *et al.* (1988), *supra*; and Miller (1990), *supra*.

Packaging cell lines capable of producing retroviral vector particles with chimeric envelope proteins may be used. Alternatively, amphotropic or xenotropic envelope proteins, such as those produced by PA317 and GPX packaging cell lines may be used to package the retroviral vectors.

In some embodiments of the invention, an antisense nucleic acid is administered which hybridizes to a gene encoding a G protein-coupled receptor of the invention or to a transcript thereof. The antisense nucleic acid can be provided as an antisense oligonucleotide (*see, e.g.*, Murayama *et al.*, *Antisense Nucleic Acid Drug Dev.* 7:109-114 (1997)). Genes encoding an antisense nucleic acid can also be provided; such genes can be introduced into cells by methods known to those of skill in the art. For example, one can introduce a gene that encodes an antisense nucleic acid in a viral vector, such as, for example, in hepatitis B virus (*see, e.g.*, Ji *et al.*, *J. Viral Hepat.* 4:167-173 (1997)), in adeno-associated virus (*see, e.g.*, Xiao *et al.*, *Brain Res.* 756:76-83 (1997)), or in other systems including, but not limited, to an HVJ (Sendai virus)-liposome gene

delivery system (see, e.g., Kaneda *et al.*, *Ann. NY Acad. Sci.* 811:299-308 (1997)), a "peptide vector" (see, e.g., Vidal *et al.*, *CR Acad. Sci III* 32:279-287 (1997)), as a gene in an episomal or plasmid vector (see, e.g., Cooper *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 94:6450-6455 (1997), Yew *et al.*, *Hum Gene Ther.* 8:575-584 (1997)), as a gene in a peptide-DNA aggregate (see, e.g., Niidome *et al.*, *J. Biol. Chem.* 272:15307-15312 (1997)), as "naked DNA" (see, e.g., U.S. Patent Nos. 5,580,859 and 5,589,466), in lipidic vector systems (see, e.g., Lee *et al.*, *Crit Rev Ther Drug Carrier Syst.* 14:173-206 (1997)), polymer coated liposomes (U.S. Patent Nos. 5,213,804 and 5,013,556), cationic liposomes (Epand *et al.*, U.S. Patent Nos. 5,283,185; 5,578,475; 5,279,833; and 5,334,761), gas filled microspheres (U.S. Patent No. 5,542,935), ligand-targeted encapsulated macromolecules (U.S. Patent Nos. 5,108,921; 5,521,291; 5,554,386; and 5,166,320).

C. Pharmaceutical Formulations

When used for pharmaceutical purposes, the vectors used for gene therapy are formulated in a suitable buffer, which can be any pharmaceutically acceptable buffer, such as phosphate buffered saline or sodium phosphate/sodium sulfate, Tris buffer, glycine buffer, sterile water, and other buffers known to the ordinarily skilled artisan such as those described by Good *et al.*, *Biochemistry* 5:467 (1966).

The compositions can additionally include a stabilizer, enhancer or other pharmaceutically acceptable carriers or vehicles. A pharmaceutically acceptable carrier can contain a physiologically acceptable compound that acts, for example, to stabilize the nucleic acids of the invention and any associated vector. A physiologically acceptable compound can include, for example, carbohydrates, such as glucose, sucrose or dextrans, antioxidants, such as ascorbic acid or glutathione, chelating agents, low molecular weight proteins or other stabilizers or excipients. Other physiologically acceptable compounds include wetting agents, emulsifying agents, dispersing agents or preservatives, which are particularly useful for preventing the growth or action of microorganisms. Various preservatives are well-known and include, for example, phenol and ascorbic acid.

Examples of carriers, stabilizers or adjuvants can be found in Remington's *Pharmaceutical Sciences*, Mack Publishing Company, Philadelphia, PA, 17th ed. (1985).

D. Administration of Formulations

The formulations of the invention can be delivered to any tissue or organ using any delivery method known to the ordinarily skilled artisan. In some embodiments

of the invention, the nucleic acids of the invention are formulated in mucosal, topical, and/or buccal formulations, particularly mucoadhesive gel and topical gel formulations. Exemplary permeation enhancing compositions, polymer matrices, and mucoadhesive gel preparations for transdermal delivery are disclosed in, e.g., U.S. Patent No. 5,346,701.

5 **E. Methods of Treatment**

The gene therapy formulations of the invention are typically administered to a cell. The cell can be provided as part of a tissue, such as an epithelial membrane, or as an isolated cell, such as in tissue culture. The cell can be provided *in vivo*, *ex vivo*, or *in vitro*.

10 The formulations can be introduced into the tissue of interest *in vivo* or *ex vivo* by a variety of methods. In some embodiments of the invention, the nucleic acids of the invention are introduced into cells by such methods as microinjection, calcium phosphate precipitation, liposome fusion, or biolistics. In further embodiments, the nucleic acids are taken up directly by the tissue of interest.

15 In some embodiments of the invention, the nucleic acids of the invention are administered *ex vivo* to cells or tissues explanted from a patient, then returned to the patient. Examples of *ex vivo* administration of therapeutic gene constructs include Nolta *et al.*, *Proc Natl. Acad. Sci. USA* 93(6):2414-9 (1996); Koc *et al.*, *Seminars in Oncology* 23 (1):46-65 (1996); Raper *et al.*, *Annals of Surgery* 223(2):116-26 (1996); Dalesandro *et 20 al.*, *J. Thorac. Cardi. Surg.* 11(2):416-22 (1996); and Makarov *et al.*, *Proc. Natl. Acad. Sci. USA* 93(1):402-6 (1996).

X. ADMINISTRATION AND PHARMACEUTICAL COMPOSITIONS

Modulators of the G protein-coupled receptors of the present invention can be administered directly to the mammalian subject for modulation of G protein-coupled receptor signaling *in vivo*. Administration is by any of the routes normally used for introducing a modulator compound into contact with the tissue to be treated and well-known to those of skill in the art. Although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

30 The pharmaceutical compositions of the invention may comprise a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular

method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention (see, e.g., Remington, *Pharmaceutical Sciences*, 17th ed. 1985)).

The modulators of the expression or activity of the G protein-coupled receptors of the invention, alone or in combination with other suitable components, can be made into aerosol formulations (*i.e.*, they can be “nebulized”) to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for administration include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, orally, nasally, topically, intravenously, intraperitoneally, or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. The modulators can also be administered as part of prepared food or drug.

The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial response in the subject over time. The dose will be determined by the efficacy of the particular modulators employed and the condition of the subject, as well as the body weight or surface area of the area to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

In determining the effective amount of the modulator to be administered a physician may evaluate circulating plasma levels of the modulator, modulator toxicity, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

For administration, the GPCR modulators of the present invention can be administered at a rate determined by the LD-50 of the modulator, and the side-effects of the inhibitor at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Table 1 below indicates, by identification in the "LifeSpan Cluster ID" column, sequences encoding putative human G protein-coupled receptors that were identified by low-stringency protein- and DNA-based blast searches of publicly available databases. "Acc. No" indicates the accession number of the sequence in the database from which the sequence of each putative receptor was identified. The type of database from which the sequence was identified and the length of the sequence in base-pairs (bp) are indicated in the "Database type" and the "Sequence Length" columns, respectively. The sequence is shown in the "Sequence" column. The column designated "LS Cluster Name and/or Representative Sequence (SEQ ID NO) provides the name of LifeSpan's gene cluster for the sequence as well as the sequence ID of another representative sequence for the cluster, if available. These representative sequences are provided in the sequence listing following Table 1. Table 1 further shows information about the closest homolog of the sequence. The name, accession number and length of the closest homolog are shown in the "Homolog Name," "Homolog Accession No." and "Len" columns, respectively. Length is given in number of amino acids unless otherwise indicated. The table also indicates the position ("From" and "To" columns) and length ("Aligned") of the region of significant identity between the sequence of interest and its closest homolog, as well as the percent identity ("Percent") over the described region.

Table 1

| LS Cluster ID: | Current (Original) LG NO. | Acc. No. | Database Type | Sequence Length | Homologous Representative Sequence (Seq ID No.) | Homolog Name | Homolog Acc. No. | Length | From To | Aligned | Percent | | |
|-------------------|---------------------------|----------|---------------|-----------------|--|------------------|------------------|--------|---------|---------|---------|-----|----|
| 30875 | AF086432 | GenBank | | 589 | ATCCATGACT GCTCAAAC TAAAGTCTTG TTGGGGCTCA AATGGGTCTAC GGCACTCACC TATGTGAAACA GTCTGCGTTG TGTCGCTGCT CGTGGTAG TGATCGATG TTACATAGCC ATATCCGATT ACATTCACAA ATTCCAGG CAATTCATAA GTCACTGCAAG CCGAAAGCCA AAACATACCC AGAGCATCAG GGTGTGTTG GCCTGGTTT TTACCTGCTT TCTACCATAT CACTTGTGCA GAATTCGCTT TACCTTGTG TACCTTGTG CACTTGTGCA GCCTTGTGCA CAAATAATCC TATATTACTG CAAGAATTTG AGCTTGTCT TGTCGCTG TAATGTGTC CTGGATTCGA TAATTTACTG TTTCATGTG AGTCATTT CNAGAAGGCT GTTCANAAA TAATTAAGA GAATCCAGG TGAAAGATC AGATCACTGC AATAGCTGAG AATAGCTGGA GTTCGCTAAT ATTAACTTAA CACTGATGIG TAGCAGTGT GTRAAATTAAT GTTCGCTT ATTGTGTTT GAAATCTATA TGTCAGT GTRAAATTAAT GTTCGCTTIC ATTATCTTAA AAAAARRAAA | GHR 87 Q15391 | ORF | | 338 bp | 172 | 332 | 160 | 41 |
| 191172 (46930) | AA758208 | Duest | | 986 | TGTAATTTAA CACTGGTATG TTATTTTCAAA AAGGGCTGGC AAGGGATC AGGGATGGAA CTTTTTACATG DGGGGTCTCA ATAGGAGGA GGAGGGTAG TGAGGAGGG TCATTAATTC TTGGCCATCT CAACACATA ACAATGCT ACCTTGACTC TGTAATCTAT TGATGACTTC TTAAAGTCA CCTGAAATAA CAAAGATTT GGTCGTGATG GGAACTCTGT AGCTCTGGG CATGCTTGCC AAAGCACTGAA GTCTCTATTC CCAACTCTT CTACGATATA CGAGCTGAC GCCAGAGAGA GTCACTAACCC AGCCACAGG AGCAGGGC CAGGAAGAA CACCAACCTT CACATAGTAT TTACACAT TGTCGCTCG GTGCTTATC TCAATGATTT TGTCGCAAC ATCTTGCTGA ATGGTTGAGC AGTGGTATC TTCTCTTCCACA TTGGCAATTA ACCAGCTGA ATTCAGCTTA TCGCTATGTC AAGGGTGTGCA AGNCATCA ATAAGCTGA CATTTCGATTT TCGCTATGTC CAGTATGCGAG ACATCTCCAS TGGTGGAC CTGTGCTACTT CAGGTTTATT GGGGGTGATA GGCTATCTGAG CGCATTCAT TAACTCCAT CGTGTGTTCT GGGACTCTTC GGATGTAAG GCATTAACCC TCCAGAAGGG TTTCATCTT TTCTGTGTTG GGGGGATCA CGCTCTGTTT GTCCCTGTC CAACATCAT TCCGACAAA TGGTCACTC ACCGGAGGA CGATTTCA TGGACTGGCC CAGACATTA AACGCTCTCTG GGGGGCTCTG AATGGTAA CGGCGCCAC CCATGGAAA CGCCCTGCTT GGTGGTGA CGCTCATGAT GATATCGAT CGCGATGATA ACTAAGGCA TATCCGGGT AAATCC | Q15391 | ORF | | 338 bp | 24 | 63 | 40 | 55 |

| LS Cluster ID: | Acc. No. | Current (Original) LG NO. | Database Type: | Sequence Length: | Sequence | LS Cluster Name | Homolog Name | From: | To: | Aligned Percent | | |
|-----------------|----------|---------------------------|----------------|------------------|---|-----------------|--|-------|------|-----------------|-----|----|
| 54602 | AA012849 | | Dbest | 447 | TCCACATCTT CACACTCCTT CAGGAACTAA AACCTDAGCC ACATTAAGCC AACGACCTCTT GATGGCGGTG GTGAGTGCC TCACTGTGT ACTTGCGCTT CATTCGGTA TGCTGATTC ACTGCATTTT GGGATAAATCTT TCACATGCGA ATGCTCTGATC TAATTAACAC GAGTATACAC CAGGCCAGCA TAATAGCCCT SEQ ID NO:5 CAGCAACTTC TGGTGGAAA GCTTGGAAA GCCTTAACCA TAATTTACA ATTATTAACCA TAACTACATG SEQ ID NO:5 TCAGTGCTT CTGTTCTTCTT TTGTTGCTCC TCAATTGTC TTTCAGTAGTGT GACATAATATCTT TCTTACTGTG GCCTCTTCTCC ATTGTGACCC AGACCAATCTT ACTCAAGCTTC GCTCAATACTC TCCCATGAG TCACATCA | Q68555 | Pheromone receptor VN6, rat | 310 | 35 | 182 | 142 | 34 |
| 55728 | A1024852 | | Dbest | 505 | TITTTTCACT TAGATAAACAT TTATTTGTT AGAGCAGATG TTGGAGACT BTI TCGTTTCAAC AGCTTAACCA TCCAAACAA CAGGGACAT TTTCGACAA protein SEQ ID NO:6 TCTGTAATAT TCCTCTTCAA TCTTCTDGA TANAGAACAC AGGATAAANA AAATGAACAT CCCTGGAA GCNTGCTGA CTGTCGAGAG CTAGAGCTGTA AGENTGTGCC SEQ ID NO:6 ACCATGATG CTCGACAC AC TGGAGAACCC CAAGAGATCC AGENTGTGCC GAGAGGAC AGAGAGGCA GGCGCTCTCC TGACAGAGAC CTTAGTGTCT CAAGGAACT AACTCTGT TTCAACCTGT CAGTGTGAGC AAAACTCTG TATATGATGA CTCCAAANG CAAGAGATGA ACAGAGATGA TTGGAGATCA TGSTCCATAAA AACCTCCAA TAAGGTGT TTGGGTGCTA AGCCPACATA CPTWGTGTG GCTTAAATPAT CGTATCTCA GTGCTGCGA AAATCCAACT ACCAC | O94867 | KIAA0768 Protein [Fragment]. | 872 | 405 | 548 | 145 | 53 |
| 160221 (121659) | T19393 | | Dbest | 379 | GCACTCTGT CCTCCCTACT CGACTTCTT CCTGACATA GGCTCTGCGAG CPR27 TCTTTTGA TCCGTTGAGA CCTCTTTTAT TCCATCTGTG GTTCCTTTT SEQ ID NO:7 AGACCATGCG CAGTTCTCA TAAAGCTT CTTCATGCAA CGTTTCTCAA ATACATTTT GCAGTGTAGT AGAGGAACTT AGGGGAACTT CTCACTAAAGT TAGATTCTT GAGGATCTC TTGGTGAAGC AGAGAAGATG GGGGGTGGG GSBAAGCTT CGGAATGSCC CTCGTTGAGT CCGGCTGAG GGTCTTGTG CTCGCTGCT TCTTGTGAG TCTCTGT | AF027955 | G protein-coupled receptor (GPR27), Mus musculus | 2679 | 2548 | 2626 | 79 | 93 |

| LS Name and Chemical Sequence | Homolog Acc. No. (SEQ ID NO.) | Homolog Name and Function | Length | From | To | Aligned Percent |
|--|--|---|--------|------|-----|--------------------|
| AATGGAAA GAACTGCGTC GCTCTGTTG ACCCTGTGAA TCAATAGAA ACTCCUCCAC ACTTGTGTC TCTGTTCTCC AGGACCCATA TTTCPTGGCC ACTTGTGAA CGTGTTTTT GAAAGATCTT CCGATTTAA CTAATGGAT TGGGTGAGG CAGCTGTGA AGAGTGCCTG GCTTCTGTTG ACTTGTGAA CGAATCCCAT GCGTTGTCG ACCTGTGAC TGTTGATCAG GGAGTAGATG ATGCTATGG CTCGGCAGA CTGACAGA CTGACAGA TTATAGGCA GTTGTAGGAC AATGAAACT ATGAGACTG TGAGAGAC TTAGAGGT CGAGATTT SEQ ID TAATGTGCG CAGCTCTG AGTGTCTTG CTGTGATCAA GTAGCACACC NO:34 CCCATAATAA GAAAGGTC TACAAATCCA AYAGAGACT CTAGATTTG AATCAATGG TTCTATGG TTCTCTGCT GTATATAAA GCGGGGAAA ATGGGATG ACCTAGCAT CTCATTTACT GTATATAAA CGGGGGGGG TAGGTCTAGC AAGATGGAG CCATCAGAC ACAGAAAG AGTATCGAG ATGGGTTTCC CACTCTGTG TGCTGGGAA CTTGATAC TGCACTAT CTGCTPATGC TGAAACAASC TCGAAACTG ATTICAGAGA CAAAGTTTG TGTTGACAG GCTGAGGTT TTTCGATC TATTTCCT AAACCCACC CATTGAACTGC ATTACAGCC CAAAAAGGCA GAGTGTATG AAGGAACTAA TCTGTCPACAG CCAAATTCAG GATGTACACA TCTGTGTT CTGTGATATG GCATAATTTG CCATCACTTCA GGAATGCTCT GCAATGCCA TGAGBAAAC TATGTGAGG AAVACAGGA GAAAACTTT TCAATTTCT CTGACATCTT CTTGTACACA GATGAGTCA TATTGACTGT AGTCATAGI GCATCATT TCATTTCTCC CATTATAAA ATGGTGTAC TGTTCTGTT CCAAGCCAT GGCTCCTAT | P32248 | C-C Chemokine receptor type 7 precursor | 378 | 21 | 156 | .39 |
| TTCCTGAACT TTCTATCAT AATTCATAG ACATATGGAT TACAGATGGA GTTGGAATT CCTATTAATT GCAAGTATG AAAATATCTC TTGATGTTGA CATCATATA TTCTTCTICA AATTTACTCT ATTCAATCAT CATTGGCA SEQ ID ACATGAAATG GIGCCAGCA CACGCAAG CTGTCACTT NO:8 CATATGACA GETCTGTTCTC TCTTCCATAA GAGGAGGAG GAAAGGGGTC ACAGGATGA AGGTGIGTA GATCTCTGG TGCAAGGGC TGTCCTCTC TTCUAGCG CAGATGTTT CCTTTCATA TAGGAGTCA TATTTGAACTC CAAGTGTG CAGTGGCAC ATGGGATC CTACGATGAC TGCCACAGC CAGAACACG CTAGCATGTT GAAAGCCCT CG | Q9YSX5 | G protein- coupled receptor | 522 | 260 | 334 | .68 |

| LS Cluster ID | Acc. No. | Current (Original) LG NO. | Database Type | Sequence Length | Sequence | LS Cluster Name and Ref ID | Homolog Acc. No. | Homolog Name | Len. | From | To | Aligned Percent | |
|-----------------|----------|---------------------------|---------------|-----------------|--|----------------------------|------------------|---|------|------|-----|-----------------|----|
| 160324 | A1090920 | Dbest | | 455 | ACCTGGCTT CCACCTTT TTGCTTGTAC TTTTCCTTA ACCACATTA AACAGAGCA TTCTATGAA AAATATACG GCAATATT ATTACATT SEQ ID NO:9 | GPR86 | Q15391 | ORF, complete cds | 338 | 84 | 230 | 43 | |
| 160435 | AA804531 | Dbest | | 599 | AACTGGAGG GCACCGCTT CGCTGGCTT GGGCCCCAG CCGGGGGCTC TGAGCACCG CCGTGACTA ACGCTTCA TCCTGTTAAC TGGAGACCT CTGAGCTTC ACCTGTACT ATGCTGGAA ACCGACGGAA AGCCACGCC CGGACAAAGC GAGCTGGCTT CCAGATGCA GTGCTGGAA ACCGGGAAV CGGGGGCC CTGCTGGTG TGACTCTGG GTCAGCATCC CGGGCAACTC CTGCTCTCTG TGGGNGCTGT GCGCGCAT GCGCACCTG AGCCCTACGS ACCCTGATGG GCGCAGCTG TTGCCPCTCC AAATCTACTA CCATGGAAC CGCCACACT GGTTAATGG GETGCTCTT TGCAACCTGG TGACCTGGC CTTTAAGCA AACATGTAGT TCAGCTNCT CAGCTGANC TGATCTAGCG TGAGGGCTC CTGGGGCTC ATGAGGGT TGAGGGCT | P2Y | P55085 | Proteinase-activated receptor 2 precursor | 397 | 62 | 172 | 111 | 38 |
| 190711 (160444) | AA883367 | Dbest | | 400 | TCTGAGGAC CTTGAGCAGA AGAGCATCA AGCAAGTGGG AAACAGACA AAACCCAG AAAGCAACT ACPTTGAGA TGAGACTCC ATAAGTCAG CTAGAGCCAT TTGAGCAGA CTGAAACACA AATGGAAAC AAATTGAGA TCTGAGGATA TCTGAGACGC AATGCTCAA CAGGAAGTAG TAAGCTTC TAGGZAGGT CTATCTTC ACTAGAAAA TGGAGATAG GAGCTCC ACCACGCTGA CTCCATTAT GAAACCCAG GAGTCACTT CGAAGGG CTGTAAGGG GCAAAATTG TGGAGCTCA TGCTATAGT | GPR85 | Q9Y5N1 | Histamine H3 receptor | 445 | 41 | 134 | 95 | 34 |

| LS Cluster ID: | Current (Original) LG NO.: | Acc. No. (160457) | Database Type | Sequence Length | Sequence | Similarity | Homolog Name | Homolog Accession No. | Homolog Length | From | To | Percent Aligned | | | |
|----------------|----------------------------|-------------------|---------------|-----------------|---|------------|--------------|-----------------------|----------------|------|----|-----------------|-------------------------------|---|--|
| | | | | | | | | | | | | | ES Cluster Name and RefSeq ID | MS-related G protein-coupled receptor MRG | |
| 191218 | AFO03828 | | Dbest | 503 | TTCGAGATCA TCACCTGCAC CCGGGTGAAT TGTTTGATCA TGTCTTGATC TGGGCCAGT CTGCCCCCTG TGCTCAGAT CCTCTGNGC TCCAGGGTC TGCCACTGAC CGACGTGAC CGGACATTG TGCTCAGAT GCTGGTCTCC CTCCCTGCGS CTGCTGCGCTTG TGCTCATGAG TGCTTCCTAA TATTATGATC CTGGEAAGGAT TGCTGATCTA TATTTTGCA TATCTGATCA GTTCTGATG TCCCCTCATC TGCTTACAGC AGGGCCACR CCATCTGCTG GGCTTCTTA GGAGGAGTG GCGGNTGAGC CACCTGATCC TCAAGTGGC TCMCAGGGS GCCTGCGAGG ACATTGCTGA GTGGATCAC AGTGAAGGAT GCFTCGTCA GGECACCCCG AGTTCAAGG AGGACATCTG GTGTAGGGAT GCAACCCTCT ACTTCATCA TATTTATG GCTTTGAGG GCAACTTGC CCC | P35410 | | 378 | | | | | | | |
| 160458 | AID64302 | | Dbest | 491 | TGTAAATAA AAGCTTATG TGGATAGCA TGTAGAACG AGTGAACAG GCATGGCCA ACAAATGAA CAACTAGAA CAGCTGACTA TAAAGCTAG CGCTGGCCA CCAAGCAAC CTCGGCCCTG CAACTATCTTG CAGTGTATG GTGGTGACTT GGAGGAAATC CCTATGTCGA GTGAGAAA AATATGATG CTGACCTCTT TATATATGA AATAATCTA CTCAGAGTC CGTCAETAG CTGGAGAAG TGATGTTGA AGTTTAAAC ATCGATCA GGTCAGT TGTCATCAA CCCATGGCTA AATAGCTGA CGGTTCTGA TCAAAAGTGA TCCATATGT GAGAACATTA ACATGCTGA AAAAGTGTCT ACAGATTATA TGENGAAAT ACCTGATGGG CTCCTGAGA GACTAGAGCA G | U45983 | | 1944 bp | | | | | | | |
| 191168 | AA274112 | (161362) | Dbest | 542 | GCCTGAGA AACCTTAACTG TTCAAGGTTT ATGTTATGG ACATGTTATA TCAGTATATC TTGACGCTTA platelet CCTGAAAGCC ACCAGGGCT TAAACCTG ADP CTGCAAGAT TCTTCTGTT CAGCCCGAC receptor CTGCTTAACTA TGATTCAC CTCATCTGG CTTAATTTCA AATGCTCTT TTCTTAAGT CCAAGGAGG CTCAGTGG TCAATTCTA CTGCAACTG ATTTCCTGA TAAATTTT AATTGTC GTTCTGTTATA GCCTCATCAC CAAGAACTC TAC'CGCTC' ATGTCAGAAC AAGGGTTCA GCCTAAGTC GCAAGAGAA GSTAAAGTC AAGGTTCTCA TCATCATTCG TGTATTCTTG ATTTCGTTTTC CAACCTCTCA GTTGTGACCG ATTCCCTACA CCCTGAGGCA GACTGSSGC | Q15391 | | 338 bp | | | | | | | |

| LS Cluster ID: | Acc. No. | Database Type: | Sequence Length: | Segment: | LS Cluster Name and Repetitive Nucleotide Sequence (SEQ ID NO.) | Homolog Name: | Homolog ID No.: | From: | To: | Aligned Percent: | | |
|----------------|----------|----------------|------------------|--|---|-------------------------------------|---------------------------------|-------|-----|------------------|-----|----|
| 162615 | AJ479284 | Dbest | 508 | TGAAACCAC TGAGACATTG TTTCTTATT CCAGAGCCC TAATACTAA | AACCGATCG AATACTGAGC ATATTCCTT CATTGACATT TGTCTCTAAA TGTCAGCTTG TTCTGGAAT TTTCATGAT TTTCATGAG TTTCATGAG CTTGCGAG GAAAGCTTT TATGAAAGAA TCCATTAATGG CCTTGTGTTT GGATGGTTT GAAAGCTTT GAGAATTTGG GTTGGCAGA TAATCTGAT GAGTTTGCT TTTCATGTTG CTTCCAGAA CTTAAGGGAG AGCACTTGT GAACAGAAAGT TTGCGAGCT TACGTGCAA GAGTTAAC AAGCTTAAC TAATAATTC CGGGATGCA TTGAGTAAAG CAALATAAC ATCCAGAAC AGATTCGGC TTGCCACTAT TGTTCTTAT CCAGCCCC AGTTGAGCC TAGAGGAGG GTCAAGATGA GGAGCTTC | 094858 | KIAA0758 protein [Fragment]. | 986 | 860 | 961 | 104 | 47 |
| 168928 | AA551068 | Dbest | 343 | GGGGTCTT AATGGGACCC TGGCAACTT AGTGGTACCT AGTAAAGACT GTTCACTTT CCACATGAA CTCCATCATG AGAAGACTGT TTCTTACTCT GTTTCTTACT CGAGATAGT TTTCATTTATA GAAAATGCG TGTCTTCAG TGCAATCAGA GTGCTTCTTT TGTCTAGGCA CCCAGAAGAA TTCTGATACT TTCACTCAC CAGCTTCTCC CCCAGACCTT CCCAGAGAA ATGGCCACT CAAGACATCC TGGCTCTCTT GACCTCTGTT GGTGTCATCT ACTGGATGAA TTTCATCATC TCATGCACTT CAACCTGCCT ATGGCAAT GAC | Y17766 | M2 pheromone receptor, Mus musculus | 959 bp | 700 | 752 | 53 | 88 | |
| 189873 | AC007104 | Genomic Clone | 852 | GCTCTCAAG CCACTGCC AGGGAGGCC CAAGTCCCTG GCCTGGGCC TGCGGGATGC ACTGGCAGCC TACCTGGCTT GGAGCAGCG CGGNCACCC TGGACGAGG TGAAGCAGAG TGGCACTTC TGGCACTTC AGCCCGGGC ATGGGCGG GCGGGCGAG CAGACAGG CAGGCGCTGA GGCTGGGG GCAGGGCGAG CAGGGAGG CAGGCGCTGA GAGGCCAGC AAGCCAGAG CAGGGAGG CAGGGAGG CAGGCGCTGA GAGGCCAGC AGGCAAGCC CAGGGAGG CAGGGAGG CAGGCGCTGA GAGGCCAGC TAGCGAGTG CGAACCCAC TGGCACTGC TGGCTCTGC TGGCCCGC CACGCTAGC GCCGGCTTGG ACCGAGGAA GGTGTCAGG AAGCCATGA CTTGCAATGC CGCGGGCC GACGGTTCGC GCGGGCTGC CAGACAGC AGCTGTAAGG GCATOTCCAG CGGGCCAG AGGGCGCTGC AGGGCGCTGC ATTCACTGG AGGGCGCTG AGGGCGCTGC AGGGCGCTGC AGGGCGCTGC CGCAAATAGG CAGGACAGT GGTTGGATA GGAGGCCAC GCGAGTAC ATCACCAAGG GACGGCCAG CAGGGCTGC CGGGGGCCA TGGCGCTGC CGCTCCAGG GCACCCGGG GTTTCATGG CTGGCTCTGC GCGGGAGCC TGGGAAGTG AGGGCGCTGC AGGGCGCTGC CGGGCGCTGC | P21918 | D(1B) dopamine receptor | 477 | 28 | 266 | 221 | 27 | |
| LG155 | | | | | | | | | | | | |

| LS Cluster ID: | LS Cluster ID: | Sequence Length: | Sequence: | Name and Description: | Homolog Name: | Homolog Acc No.: | Percent Identity: | | |
|----------------|----------------|------------------|--|-----------------------------------|---------------|------------------|-------------------|-----|-----|
| 189874 | AC008571 | 813 | TCTTAGTGTTC TCTCTGAACT TTTCAGGCTCTT CAAANAAAAGA CGCGATGCTGC CTCAGGCCAC CCAGGCCATC ATATCTACTC TGAGTGCCAT GAGCTGATAG AGGACACAGCA TGACAGCTAT GGGGGAGGAGG TAGAATAGGA AGGAGGTRAC CTGGATGATG AAATTTGAGA TCCACATGGG CTTCATGACCC GTACAGCTGG CCGAACCTGGG GACCAGGGAC CCATATTGGGA AGTATGTTGAA CTTCATGGCA TGATGCTGG TGTGGCAG GGAGAGAGC ACAGGAGAC CCAGAGCAT GCCGAAGGATC CTTGAGGAGC GCCTCTGAGT GCTCTGAGT TTGGCGCGA ACGGCTTAG GATGGCCACG TAGGGCTCCA CGTGTGACGT GETGTGACTCTG AGGATGGGG CGAAGGCACAC GGTCTCAAAAG AGGGCCCTCT TAAGATGICA GCCCAACGGG CCGAACAGAA AAGGGTAGTT GCGCACATC TCTAGACCT CCAGGGGGAT TCCAGAGGC AGGAGCAGGA GGTCAAGAAC CCCAGGCTG AAGGAGCTAT AGTGGCTGG CGTGTGACATA GCCTGGTCTCT GCGAATCAC CTGGCAGACCC AGGAGCTTGC CAATGACCCC CAGCAAAA ATTGGGCACAT ACACCAACGA CACGGGAGG AAGGAGTGGC TCGCCGAGG TCAGCAGAGG AAGGCAGAT ACTCTCTGGT GCTGTCTGAG TGTTCAGG AATGGATCTC TAGTTCTGC TGGTGTGATCC AGGAGGATT CTCAGTTT TCCATCCTG ACATTAAAT CCA | Orphan G protein-coupled receptor | O43664 | 403 | 17 | 235 | 218 |
| LC349 | | | | | | | 56 | | |

| LS Cluster ID: | LS Cluster Name: | Sequence Length: | Database Type: | Sequence Range: | From: | To: | Aligned: | Percent: |
|---------------------------|------------------|------------------|----------------|---|-------|-----|----------|----------|
| Acc. No. | Repres. Ac. No. | Homolog Ac. No. | Homolog Name: | Homolog Acc. No. | Len: | | | |
| Current (Original) LG NO. | | | | | | | | |
| 189876 | AP000898 | 1113 | Genomic Clone | GGCCAGGGC TCGGAGGCC CGGAAGGCC TCGGGGGC TGTCAAGCCC | 325 | 7 | 306 | 297 |
| LG1543 | | | | CCATCCATT GTTCGCCAC GCGGCGTGT GCGACACCTT CCCACACTT CCGCTGGGC TCCTCCGAA GCGCGTGTG GAGACAGTC CCAGGACCC CGCGACGGC TGTTGGCAC SEQ ID CTCGGAGTC CGCGAGTC CCACAGGAA GTAGATGAGC CAGGACCTC NO:15 TGCTGTTAC GGACGAGAG AGCGTGTAC CCGAGGAGA AGCTGAGCA CAGGACCTC ATCTCGGG GCAGCTCAA CGAGTAGAG AGCGTGTAC CCGAGGAGA AGCTGAGCA AGGCAGGAA CAGAAGGAA ACAGCAGAAC CGCGACGGC AGAGGCTAGG ACCACACGA ACAGCGGTG GGGCTCCGC CGCGACTGT CGCGACTGT AGGGGCTCT CGGACCCAT ACAAAGAGGG TCAGSCTGGA CAGAGTCATC ATCGGGGTTA AGACCCCAT GATGAGGGG GCCM9GACCA TGTGACCTT GAGGACCCA TGTGATGGA ATTTGAGAA CTGTGCGCG ACGGAGGG TCAACCGGT CMTCAAGGAGA CAGAGGTC ACAGCAGGC ACAGCAGGC GETGACAGT GCTGGGGCG GTGACCTTG ACGCGATAG GGAGAGGAC AGAGAGACAG CGCTGGGTGC TGATGCGGT CAGCGAGTC AGGCCCCATG TGTAGGZAA CTACATCGT CTCTTCATCA GCTCTGGAC CTTCTCACTG GPATTGACCA GGGGCTGGT TTCCAGGCTG AGCGTGAAG CTCAGTGA GAGGAGGG AGTCGCGCTG CGCGCCGGTT GAGGTATAAG ATGGAGAAGG GCTTCCTGTG CATTGCAAAG CCAGGAGCC AGATGACCAT GCTTGCTGT GCACTCCGGC ACAGGAGGT GNAACATGCC AGGGAGCTCA GCACAGCTTA GCGCTGTGCG ACTGTCCTCC CTCTGGAAATA GTTAAAGGT GACTCGAGG TCCACAGCT ATTCAAGTGC TGGTTCATCC CTACAGGAGG MACTGTACG AATGTTGAAT TGTGTGTGCG TGGACGAG GGGACCCCT GGTGACCTT CGATTTCGA GCTTCAGAGC TCTCCCTCC AGG | 325 | 7 | 306 | 297 |
| 189878 | AC016362 | 504 | Genomic Clone | CCTGGGAGGT CGGTGTTCC GATCTGGCA CAGCAGGAGG GTTGAGCTGG TCTTTTTAAA GTGGCGTGTG CAGCAGGAGT AGGAGGGGG GTTGAGCTGG GCAGGCTCTGG CAGAAGCTGT TCACTGGAC CATGAGCTT TAGGGCTCC | 479 | 369 | 479 | 111 |
| LG1143 | | | | CGGTGGGGAT GAAACCTAAC ANAATGGCA AGTTCGGCTG TGGACTTGG CGCTCCGGAC CGCGATCTGG CGACCTCTGG TGGAGGGAT GCTTGGGAACT TTGGGGCCA CGTGGGGCGC AGCGCATGC CAGNCGCGT GGGAGGACA ATCTGAGGC TGGACACAC TGTGGGCTA GCTGGCTTG TCRAATTGG TGTCTGGAC CAGTGGGAG SCITGGGAG AGCGCCCCGG CTCGGAGCTG CGGGATGAA TCCAGGCTT ACTCTANAGG ATCCCCCT CTCC | 479 | 369 | 479 | 111 |
| | | | | P08173 | | | | 92 |
| | | | | Muscarinic acetylcholine receptor M4 | | | | |

| LS Cluster ID | Accession No. | Database Type | Sequence Length | Sequence | LS Cluster ID | Name and Representative Sequence | Homolog Accession No. | Protein ID (SEQ ID NO.) | Aligned Pairs | Aligned Pairs | Percent Identity |
|---------------|------------------|---------------|-----------------|---|---------------|--|-----------------------|-------------------------|---------------|---------------|------------------|
| LG1390 | AI049739 | Genomic Clone | 330 | CCTCCATCA CCTCAAGAAAG CCCCTGAAA AGAGACCCC TGTGCTTTT CCGTAGAGT TACTTGCA AGAGTAGAT GACTAAGTC GGCATGCTGA CCACATGGG CAGTGCAAC AGAGGTGAA AGCAGAGGT CAGAACCTT CCTGCCAGGC ACACAGGAC AGCTAAGGG AGACACCAA GAGAACGGC CTATGATCA AAGCAGACA ATGATAGGT CCTGATGGG GAGAACCTT GGGAAGAGCA CAGACAGCA ATGATAGGT TTAGTACAGC CTTGCCATT ACGAGCAAA TTAGTACAGC AGATGCCATA ACTGCAVACA ACATGGCAGG CACTTCCTCC AGTACTACAA GACTCTGGTA ACTGCAVACA AAGAAGGG CAGGGCCAC TCACATGG CAGAGGCTG CTTGAGTGT CGCTGGGG TGCCAGRACC AGGTEGGACA GAGGCCACAG CRATACTCAT SEQ ID GGCCACCAAGG AGACAGAAC CCACCTGTC GAGAAATAG GAGACAGAT NO:16 CCAGAAGAC ACCACCTGC ANTGCCTGC AGTACTACAA GAGAACCTC TTATGAGT TGCTCCAGCA GGATCCTGCT TACAGGGAG AGGTGACCA TTATCAAGT GGCCAGGTTA AGGATCTAGG TCATATAGGG GCCTCTCCAG ACCTGTGAGT AGAGAAGCCA GCGATCACCA TCATGCTCTA CGAGTCCACA GGAGGCCACC AGACATGCA GGGAAAGAC CACCTGGCTG TCACCAACC ACTCACCTCC GGTAAGGCTC AAGTGTGAGC GTCTGGGGC CTCAGGTC TT TGAGGTTACC TGCCCTCTGA TGGCAGGAC TGGAGAACTA TTATCCGGGC Putative TGCTCTGGT GCTGCTGCC TGGTGGCTT CGTGGAAAC CIGTGTGTA GalR4 TGCGCACTCT CCTTCACAAAT GCTTGAAAG GAAAGCCATC CATGATCAC receptor TGCTCTGGT TGAACTCAG CCTGCTGAT CTCCTCC TCCTGTGCT TGCCATCTC CGAGCTACCA CTCAGTCTG GAGTGTGAGT GATCTGAGCT SEQ ID GTTGTTGCTG CAAGTCCTCT GACTGTGTTA TCACACATG CTCGCGGCC NO:1 AAGAGCCCTGA CAATGCTCTT GTCGCCCCAA GTCAGCTCA TTATGCAA TGCCCAACC CAGCTAGTG TTTCACACT ACCCTATG GTATGCGG TGCGCTTTT GACTGGGGCT TACCCCTTA TGAGTGTGCC CCTTCCTCCAC ACTGCCAGC CTCAGGCAAA GGACCCAGT SEQ ID CCCATCTTC TTTCAGACCC TGCTCACTA CCTCACGCC ACCATGAG TGTTGAGCT GATTCCTCAC TTCCGATGTT CTGGGGTAG CGTCCCTGGCG NO:17 CAGGGGAGC ACTTGAGCT GGGTGTCGA TTGATTCCA ACTTCGATC CCACCCGGG GGCCAGGCT AGTCCCTGAA CAGTAAACAC NTCTATGGC TGTTGAGAA CTGTAAGGCC AGTCCCTGAA TGTTTCTC TAATAGCTAG CACTTCAGAC TTATCCCTGA ACCATAATCC TGCTTCTC GAAGGGGCA GTCCTGGTC AGCTGGGACA GTGCGCTGCA GCGCAGGGCA TTCTGGTC CTGCTGTC CA CTCTGCTCAT GGCGCTGGTC CTGACTGTC | P35410 | MAS-related G protein-coupled receptor MRG | 378 | 232 | 342 | 110 | 39 |
| LG1391 | AI049739 | Genomic Clone | 492 | CTTGCCATT ACGAGCAAA TTAGTACAGC AGATGCCATA ACTGCAVACA ACATGGCAGG CACTTCCTCC AGTACTACAA GACTCTGGTA ACTGCAVACA AAGAAGGG CAGGGCCAC TCACATGG CAGAGGCTG CTTGAGTGT CGCTGGGG TGCCAGRACC AGGTEGGACA GAGGCCACAG CRATACTCAT SEQ ID GGCCACCAAGG AGACAGAAC CCACCTGTC GAGAAATAG GAGACAGAT NO:16 CCAGAAGAC ACCACCTGC ANTGCCTGC AGTACTACAA GAGAACCTC TTATGAGT TGCTCCAGCA GGATCCTGCT TACAGGGAG AGGTGACCA TTATCAAGT GGCCAGGTTA AGGATCTAGG TCATATAGGG GCCTCTCCAG ACCTGTGAGT AGAGAAGCCA GCGATCACCA TCATGCTCTA CGAGTCCACA GGAGGCCACC AGACATGCA GGGAAAGAC CACCTGGCTG TCACCAACC ACTCACCTCC GGTAAGGCTC AAGTGTGAGC GTCTGGGGC CTCAGGTC TT TGAGGTTACC TGCCCTCTGA TGGCAGGAC TGGAGAACTA TTATCCGGGC Putative TGCTCTGGT GCTGCTGCC TGGTGGCTT CGTGGAAAC CIGTGTGTA GalR4 TGCGCACTCT CCTTCACAAAT GCTTGAAAG GAAAGCCATC CATGATCAC receptor TGCTCTGGT TGAACTCAG CCTGCTGAT CTCCTCC TCCTGTGCT TGCCATCTC CGAGCTACCA CTCAGTCTG GAGTGTGAGT GATCTGAGCT SEQ ID GTTGTTGCTG CAAGTCCTCT GACTGTGTTA TCACACATG CTCGCGGCC NO:1 AAGAGCCCTGA CAATGCTCTT GTCGCCCCAA GTCAGCTCA TTATGCAA TGCCCAACC CAGCTAGTG TTTCACACT ACCCTATG GTATGCGG TGCGCTTTT GACTGGGGCT TACCCCTTA TGAGTGTGCC CCTTCCTCCAC ACTGCCAGC CTCAGGCAAA GGACCCAGT SEQ ID CCCATCTTC TTTCAGACCC TGCTCACTA CCTCACGCC ACCATGAG TGTTGAGCT GATTCCTCAC TTCCGATGTT CTGGGGTAG CGTCCCTGGCG NO:17 CAGGGGAGC ACTTGAGCT GGGTGTCGA TTGATTCCA ACTTCGATC CCACCCGGG GGCCAGGCT AGTCCCTGAA CAGTAAACAC NTCTATGGC TGTTGAGAA CTGTAAGGCC AGTCCCTGAA TGTTTCTC TAATAGCTAG CACTTCAGAC TTATCCCTGA ACCATAATCC TGCTTCTC GAAGGGGCA TTCTGGTC AGCTGGGACA GTGCGCTGCA GCGCAGGGCA TTCTGGTC CTGACTGTC CA CTCTGCTCAT GGCGCTGGTC CTGACTGTC | P35410 | MAS-related G protein-coupled receptor MRG | 378 | 54 | 221 | 164 | 39 |
| LG610 | AC011386 (89882) | Genomic Clone | 429 | CTTGCCATT ACGAGCAAA TTAGTACAGC AGATGCCATA ACTGCAVACA ACATGGCAGG CACTTCCTCC AGTACTACAA GACTCTGGTA ACTGCAVACA AAGAAGGG CAGGGCCAC TCACATGG CAGAGGCTG CTTGAGTGT CGCTGGGG TGCCAGRACC AGGTEGGACA GAGGCCACAG CRATACTCAT SEQ ID GGCCACCAAGG AGACAGAAC CCACCTGTC GAGAAATAG GAGACAGAT NO:16 CCAGAAGAC ACCACCTGC ANTGCCTGC AGTACTACAA GAGAACCTC TTATGAGT TGCTCCAGCA GGATCCTGCT TACAGGGAG AGGTGACCA TTATCAAGT GGCCAGGTTA AGGATCTAGG TCATATAGGG GCCTCTCCAG ACCTGTGAGT AGAGAAGCCA GCGATCACCA TCATGCTCTA CGAGTCCACA GGAGGCCACC AGACATGCA GGGAAAGAC CACCTGGCTG TCACCAACC ACTCACCTCC GGTAAGGCTC AAGTGTGAGC GTCTGGGGC CTCAGGTC TT TGAGGTTACC TGCCCTCTGA TGGCAGGAC TGGAGAACTA TTATCCGGGC Putative TGCTCTGGT GCTGCTGCC TGGTGGCTT CGTGGAAAC CIGTGTGTA GalR4 TGCGCACTCT CCTTCACAAAT GCTTGAAAG GAAAGCCATC CATGATCAC receptor TGCTCTGGT TGAACTCAG CCTGCTGAT CTCCTCC TCCTGTGCT TGCCATCTC CGAGCTACCA CTCAGTCTG GAGTGTGAGT GATCTGAGCT SEQ ID GTTGTTGCTG CAAGTCCTCT GACTGTGTTA TCACACATG CTCGCGGCC NO:1 AAGAGCCCTGA CAATGCTCTT GTCGCCCCAA GTCAGCTCA TTATGCAA TGCCCAACC CAGCTAGTG TTTCACACT ACCCTATG GTATGCGG TGCGCTTTT GACTGGGGCT TACCCCTTA TGAGTGTGCC CCTTCCTCCAC ACTGCCAGC CTCAGGCAAA GGACCCAGT SEQ ID CCCATCTTC TTTCAGACCC TGCTCACTA CCTCACGCC ACCATGAG TGTTGAGCT GATTCCTCAC TTCCGATGTT CTGGGGTAG CGTCCCTGGCG NO:17 CAGGGGAGC ACTTGAGCT GGGTGTCGA TTGATTCCA ACTTCGATC CCACCCGGG GGCCAGGCT AGTCCCTGAA CAGTAAACAC NTCTATGGC TGTTGAGAA CTGTAAGGCC AGTCCCTGAA TGTTTCTC TAATAGCTAG CACTTCAGAC TTATCCCTGA ACCATAATCC TGCTTCTC GAAGGGGCA TTCTGGTC AGCTGGGACA GTGCGCTGCA GCGCAGGGCA TTCTGGTC CTGACTGTC CA CTCTGCTCAT GGCGCTGGTC CTGACTGTC | P47211 | Galanin receptor type 1 | 349 | 46 | 130 | 85 | 40 |
| LG455 | AC009763 | Genomic Clone | 432 | CTTGCCATT ACGAGCAAA TTAGTACAGC AGATGCCATA ACTGCAVACA ACATGGCAGG CACTTCCTCC AGTACTACAA GACTCTGGTA ACTGCAVACA AAGAAGGG CAGGGCCAC TCACATGG CAGAGGCTG CTTGAGTGT CGCTGGGG TGCCAGRACC AGGTEGGACA GAGGCCACAG CRATACTCAT SEQ ID GGCCACCAAGG AGACAGAAC CCACCTGTC GAGAAATAG GAGACAGAT NO:16 CCAGAAGAC ACCACCTGC ANTGCCTGC AGTACTACAA GAGAACCTC TTATGAGT TGCTCCAGCA GGATCCTGCT TACAGGGAG AGGTGACCA TTATCAAGT GGCCAGGTTA AGGATCTAGG TCATATAGGG GCCTCTCCAG ACCTGTGAGT AGAGAAGCCA GCGATCACCA TCATGCTCTA CGAGTCCACA GGAGGCCACC AGACATGCA GGGAAAGAC CACCTGGCTG TCACCAACC ACTCACCTCC GGTAAGGCTC AAGTGTGAGC GTCTGGGGC CTCAGGTC TT TGAGGTTACC TGCCCTCTGA TGGCAGGAC TGGAGAACTA TTATCCGGGC Putative TGCTCTGGT GCTGCTGCC TGGTGGCTT CGTGGAAAC CIGTGTGTA GalR4 TGCGCACTCT CCTTCACAAAT GCTTGAAAG GAAAGCCATC CATGATCAC receptor TGCTCTGGT TGAACTCAG CCTGCTGAT CTCCTCC TCCTGTGCT TGCCATCTC CGAGCTACCA CTCAGTCTG GAGTGTGAGT GATCTGAGCT SEQ ID GTTGTTGCTG CAAGTCCTCT GACTGTGTTA TCACACATG CTCGCGGCC NO:1 AAGAGCCCTGA CAATGCTCTT GTCGCCCCAA GTCAGCTCA TTATGCAA TGCCCAACC CAGCTAGTG TTTCACACT ACCCTATG GTATGCGG TGCGCTTTT GACTGGGGCT TACCCCTTA TGAGTGTGCC CCTTCCTCCAC ACTGCCAGC CTCAGGCAAA GGACCCAGT SEQ ID CCCATCTTC TTTCAGACCC TGCTCACTA CCTCACGCC ACCATGAG TGTTGAGCT GATTCCTCAC TTCCGATGTT CTGGGGTAG CGTCCCTGGCG NO:17 CAGGGGAGC ACTTGAGCT GGGTGTCGA TTGATTCCA ACTTCGATC CCACCCGGG GGCCAGGCT AGTCCCTGAA CAGTAAACAC NTCTATGGC TGTTGAGAA CTGTAAGGCC AGTCCCTGAA TGTTTCTC TAATAGCTAG CACTTCAGAC TTATCCCTGA ACCATAATCC TGCTTCTC GAAGGGGCA TTCTGGTC AGCTGGGACA GTGCGCTGCA GCGCAGGGCA TTCTGGTC CTGACTGTC CA CTCTGCTCAT GGCGCTGGTC CTGACTGTC | P41180 | Extracellular calcium-sensing receptor precursor | 1078 | 164 | 308 | 144 | 25 |

| LS Cluster ID | Ac. No. | Current (Original) LG NO. | Database Type | Sequence Length | Sequence | Homolog Name | Homolog Acc. No. | Sequence (SEQ. ID) | From | To | Aligned | Percent |
|---------------|----------|---------------------------|---------------|-----------------|---|--------------|------------------|--------------------|------|-----|---------|---------|
| 189884 | AC011375 | LG608 | Genomic | 1137 | ATATCAGGCC TTTCRAGTTT TCCCTTGCC AGAGGAGGA GAGCTGGTT TCTCTTTTC TGGGTGGAT CTGGGGGATT CTCAGATCG AACCTGTTCA CGAAAGCCCTT CTGACTTGC AGCTGGTT TCCTGAGACT CGTAGACAGT TGAGGTTTTT TTGGTTATCA TCCATTTCGA TACACCTTT AACGCTTCCC TGAACTCTTC CGACATICA AGRAAATAA GAGGATTTC TGAAAGATG GAAACATCA AGACTGAGA CAGGGCTATG AAACCTTGTG GTGGGGCCGG GCCTGAGCC TTCAAGATGCC ATTACCAAGC CCAGCTTAC CATTCGGGGA GCCACAGAG AGCACAGATG ATGCAATTC TGAGCATG CRCTGTAAC TGCTTGGC GTATTCGGT TCTPAGATT TGAGCTTAG TTCTCTTGT TTCTCTTGT TTTACTTGG TCTATAGCTC TCCGAATAA AACCTGGCA AAAATAATG GAAGGCCAAA TGCCAGGAGT GGTTAGAGT TACCAAATG CGACATAAAC TCTTCAGCCA CAGCTGGTC ATCCAGGAG CAGCTTCCA CRCTCTGATG ATGCCATGATG GTGCTTAAGA ACCATTCCG CAGGGTAAAC CGCTTACCA CAGTCAGAT GGCCCCCAGG ACTAACCCAA TGTTGTTGTT CTGGATACTC ACTTGTCTGG CTGGCTCACT TGCTTACATG AACATACATG TGCCACAC AACGATGTC AGGCCTTGG CTGCAATGCA TGTTGTTAA AACACTTTCG AGGACTGCA GACAACCAG CCTPAGATCCC AACACTTTCG GAGATGAGC GTAGCTGGA TAGGTGAGA AAAGAGGAG AGGAGGAGT CGCCAGGCT GAGATCAGA ATCAGGAGT GGATCATGGA TGCTTCTCTT TCCCAAGCAT TGTGAAGAG GATGGCAATC ACACAGGTT TCCACAGGA GCCACACGG CAGACGCCA CCAAGAGAGC CGGGATGATG GTCCTCCAGT CGTGGGATC AGAGGGAGG TACCTCCGG CAAGTGGAG GTAGCAAG CTCGGAG TGTCTGCGA GTTAAGCTCTT GCAAGGGCG CTCGGAG | Q9NS07 | GPR SALPR | 469 | 81 | 444 | 326 | 28 |
| | | | | | | | | | | | | |

| L.S. Cluster ID: | Acc. No. (Original) LG NO. | Database Type: | Sequence Length: | Sequence | Scoring: | IS Cluster Name and Reference Sequence (SEQ ID NO.) | Homolog Acc. No. (SEQ ID NO.) | Homolog Name | Len. | To: | From: | Aligned: | Percent: | | | |
|------------------|----------------------------|----------------|------------------|---|----------|---|-------------------------------|--|------|-----|-------|----------|----------|--|--|--|
| 189885 | AC011402 | Genomic Clone | 963 | TCTGCTCTCA AGCAGTCCTT CACACATGA CTTCATCCG CCTTGCTCTG AAGCTTGGCT GCGATGGCC TTGTAGTTGA TTCCACATGT CGTGGCTGA CTTGACAGTT CAACAGCTTG TCACTTCACG TCTGCTCTCA GTAGCTCTG CTTCATCCG CCTTGCTCTG CCTTGCTCTG TTGTAGTTGA ATAGCATCA GTCTATTG AACTTCTGG CCAGCTGCTG TGTAACTCA AGTACATCA GTCTATTG ACTTGTCATC CTGTGCCAA TTGTGCTCTG GAACAATTC ACACAGCAA AGGCAACTCT GGECCCTCTG AGGACCTGGC TTGTGGCAT TTGTGCTCT GTTCTCTACT TGATCTCA AGGAACTCTG GTGGAAGT GTCACCCCT TTGTACRAAC AGTATGATCT GCNGAATGAA ACTCAGGAA GTCACCAACT TTGGAAAGAC ATPATCATC CATGGCAAA AACGGTGTCA ACHACGCC ACTTTTCTT TGCTCTCTT CTGGCTCTG CTGGCTACTAC TGGEACTAC ATTCCTTCTG CCCTTGAGTT AGAGAAGAG CAGCTGAGTA AGTTAGCTG ACCCCTCTCG GTCTCTTGGA CTGTGGTAC CACCTCTTC CTCTGCTGGT TGCCCTTGCA AGTGWCCCTG TSGCTGGACT TOACATCAI TCGGAGAGAC AGAGAGGGGC TGAGACCAGT GCTCTTACTC AGTCTCTAC GTCTGTTG TATGGCTT ATGACAGT GTCTCTACTC AGTCTCTAC GTCTGTTG GGCTGATT CTGGAGAGAC TGTCTCTACTC TGCTGCTGAG TGCCCTAGAA CGGGCACTA GTGGAGGC AGATAGTGC TGAACTCAG CTCCCAGGCA GATAGTCTT TWA | 56 | P21462 | P21462 | TMLT-LEU-PHE receptor | 350 | 26 | 62 | 37 | | | | |
| 189886 | AC016189 | Genomic Clone | 330 | GGGCTTACCT TGATGCCCTG TGTGAGCTTG GACCTTACCA CAGCTGTGG CTGGCTCCAC CTGGCTCCAC GGGCTGCGC GCGCAGCTGG TCTGGCTCTC CAGCTGAGAC TGTGAGCTTG GCGCCTGCTC TTGTAGCTCA TGACCAACCC TGTGAGCTTG GCGCCTGCTC CTGGCTGAGCT GACTCTGGC AACCTGGCTC GCGCAGCTGG TGTGCTCTG CGCTCTGGC CCTCTGCTC CTGGCTGGCT AGATCACTT GGAGCTGGT CAGCTGAGCT GCGCAGCTGG AAGATCACTT GGAGCTGGT CAGCTGAGCT GCGCAGCTGG | 56 | P32249 | P32249 | EBV-induced G protein-coupled receptor 2 | 361 | 118 | 227 | 110 | 40 | | | |

| LS Cluster ID: | Accession No. | Database Type: | Sequence Length: | Sequence: | From: | To: | Aligned: | Percent: | |
|------------------------------|---------------|----------------|------------------|--|-------|-----|----------|----------|----|
| D: Current (Original) LG NO. | | | | | | | | | |
| 189887 LG626 | AC011457 | Genomic Clone | 537 | GATGCCATCA TGTAAACCGAG TACTGGAGT CAAGATGGAG ACATACCTCCA ACCGAGATA C TCGAGAACAC CAACATGCTG AAGCTGAGAC ATTGTGGTTTC ATGGAAAGATG TCGGGCGGT TACGTCCAGA GAACTGTAGAA AACMGTGATA CGACCCCGAT TACACGTGGAT GATGATTATA CCAGCAGTAT GGTTGTGTGT GAGGGAGGA AGGAGAAGTTT TCCAATCTAGA "TCAGAGATGAT GTGCACTTGG ATAGGGGGC AAAGGTGACA ACAGAATGAG ATTATAGGGT ACCCAACAGG ACCTTGGTC CTCGGTGGCC TTGTAGCTTT CTCATCTGC TAGCTGGCC TGTAGCTTG AGGCGCTAGA CTCAGTGAAG AGTTTGGCT GAAAGAGAG ACACGCCAC TGTGAACTGA ACTGAAAGG TGCGCTCG CGAGGCCAG GRGAGGCCA AGATGGTGT AGAATGGAGC AGGGCTTGTGA TCAAGCCCCAG CAGAGCTTG CGTGTCTGT CTGGGGTTC CTCCTACAC GAGGGTCA C GAGACATCG CATGGCTCATG CCTTGATACCA GCCTCATTCAC GCTTATGATG GCGAGCAGT GCCTACAGG GTTGATAACG CTCCTCCCTT CCCATCCAC CAGCATTCAC CCTCCCTCAG GTATTCCTCG TGATGGTGT AGAGCCCAAG GGTAACCTT TCCTACAGC CCTTGATCTCC GCGAGCTTGG AGCAGCATGA GGGGAGAAC TGCGCTTT GCGAGCTTGGT GAAAGACAA GAGGGTGTG TGGCCACAA CGAGGGCTG GCGACATCC AGAAAAGGT GGCCAGGTAG AGGAAATGAC AGAGGAGGC GGCAAGAGGC AGAGGGGGCT TCGGGCCCT GGAGAGAGGA ATGGGGCCC CAGGAGCAA GTGTCCTGG CCGAGAGGA GAGACCCATG TTGAGGAGG CGGGTGGGG GAAATAGGAG ATCTGTTC GCACCAACAC TCTCACACCC AGCAGCTACA CACCCAGGA CACAGAGAC GCGAGTATG AGACCTCAA AGCAGCTGAA CTCAGACACG CCAGAGGG TTCTTCCTGGA AGCTGTGGT GGGAGCTGG GAGGGAGAG GCAGTGAGGT GCYGGAGAG GCAGTGAGCA CTTGGGCTGG CACCTGGCAC CTGTCCTGG CACCCCTCTT TGAGCAACC CCCCCTGCCC TGAGAGAGAC TGATGATCCA GAAGACACAG TGAGGGAAC CATCTGTGT CCADAGTCC GACAGGAGCC AGGGAGGGAG TGGATAGAG TGCTCCAGT GGAAACCCC AGCCCTGTC CATACTTGA GGGCAGAGG AGGGAGGGAG TGCTCCAGT TTGCGACAC CAGCTAGTA ATACTTATT CAGTTCCATT AGCAGCAAT GGGGCCAGTG ATGCTGTGG ANTCTGAGCC TGTAGTGGGG GCGGAGATG CTGTGTCAG CAGGAAACCT GGGTCACAC AGCTGGCT GCAAGCAGC ATGGCTTARG CTGAGGGCGA AGGGGGTGGC CTCGGGGCAC AGGCTGGCAT CGAGGGTTC CACAGCAGC AGGGAGGGAG AGCTGGCCA GGGCTTCGG GCTTGGGACCA GGGTCACAG AGGGCTGGTC TCCATATATA GGACCTPTGTC TGTTGGCAATC AGGAGATTC GCAGCAC | 266 | 39 | 210 | 179 | 31 |
| 189888 LG5533 | AC010896 | Genomic Clone | 1317 | KIAA0758 Calcium receptor/CaR protein [fragment]. | 986 | 431 | 874 | 439 | |

| LS Cluster ID: | LS Cluster ID: | Database Type: | Sequence Length: | Sequence: | LS Cluster Name and Reference ID: | Homolog Acc No.: | Homolog Name: | Percent Aligned: | From: | To: | | |
|---------------------------|----------------|----------------|------------------|--|-----------------------------------|------------------|--|------------------|-------|-----|-----|----|
| 189889 | AC016856 | Genomic Clone | 420 | CATTTGGCTT TTGTGTCTT TTAGGAGGAG ATGGTGGATT TGGAAGAAA GGGGGGGGGG ATAGGGAGAC ATAGGATGACA AAGACTCTCA TGGCCTAGGT GTCTTGTCTG TGCGCTCTCA AGGGTGGGG TATGAGGAC ACCAGAGAC AGGGAGAC AGGGAGAC AGGGTGGGG CTTTGCTTCAGA TTGAGGTGGT CAGGGTGGG GCAGGTGGG AGGGTGGGG GCAAGCTGAC CAGATCTGAC AAACTAAGT AGGGCAAGT AGGGCAAGT TCCRAAGGGC CTTATACACA CTTAACAGTC ACCTGGGGC ANGGTGTGT AGGGTGGGGC GGCAATCTGG TCACCAAGAG TCACACAGAG GCTGCCTGGG TCAAGGGAAACA | AADI4370 | AAID14370 | Calcium receptor(CaR protein (fragment) | 206 | 89 | 230 | 140 | 30 |
| 189890 | AC016856 | Genomic Clone | 612 | CCAGATGACA AGGGTGGAA GAACTTGTG GAACTTGTG GAACTTGTG CAGGCGAGAA ATGGTGGCTT CTGGCTCTTT CCCCGAGATC ACCCTGAT CTCAGCGCT CGAATTTTCC AACCTTTC AGTTCCTTCC CTTCCCTCGC ACAGTGGCCA CGAGGAGGCC CCGAGGAGGCC AGCTCTGTG TTCAGTTTAC TGACCTGGAG TAGGGCTGGT CAGTTATGAC AGCAAGCAT TTGAGTGGCT GGATGAGCGAG CTGAGAAGAG AGGTGGGGG CGAGGGTGGC TGGATGGCT TCTCAAGAGT CAGGAGTGA GATGAGACAA TCAACAGCAT GGCCAGCTC ATGGCTTAAA CCGCGCTCG CAGGGTGTAGT GTCGTGTACT GCTACATT TTCTTCTGAA CTCCGGGG GGCCTCTCA AGGAGACAT GTTGATGGGA GGATGGAT CTTCCTCACAT ATPAACCCCTC GGTGCTGGT CCAGAAAGCC AGGGTGGCT GAGGGGGGGC TTGAGCTGAG CCATACACT TAGATAATA CCTGGCTTAA AGGACTTCTT GTTGGGGCT GCCGCCAAC GT | P41180 | P41180 | Extracellular calcium-sensing receptor precursor | 1078 | 136 | 341 | 204 | 25 |
| 160833 (189891) LQ5616 | AC011638 | Genomic Clone | 419 | GGAGGTTGTTTCTTGTGTTTCTTGTGTTTCTTGTGTTTCTTGTG TATTTTTAA TTCTTATT TTCTTATT TTCTTGTGTTTCTTGTGTTTCTTGTG ATGGGCTCA AGGGGCTTC AGGGGCTTC AGGGGCTTC AGGGGCTTC CAGGGAGGG GGGATCTCTG GCGGAGGG AGGGAGGG AGGGAGGG GAGGGGGGAA ATTCTGTTGGA ATTCTGTTGGA ATTCTGTTGGA ATTCTGTTGGA TCTCTGAGGA ACTCTCAGC CCACCTCTG CCGCTGAGCT CCGCTGAGCT TGGAAAGACA TATTTCTGAG GCTTGGCTAC TCCAATCTT TCTTCACCC CTGATTTAC AGGGTGTGTT AACAGAACT ACAACATGC CTTCAGAGGC CTCTTGTACTA AGGGAGAT | P47898 | P47898 | 5-hydroxy-tryptamine 5A receptor | 357 | 247 | 340 | 95 | 74 |

| ES Cluster ID. | Database Type | Sequence Length | Sequence | Cluster Name and Reproducible | Homolog Name | Homolog Acc No. | Sequence SEQ ID NO. | From To | Aligned | Parent | |
|-----------------|---------------|-----------------|---|-------------------------------|------------------|-----------------|---------------------|---------|---------|----------|--|
| 189892 LG606 | Genomic Clone | 934 | GACATGTTT COATGATCAT CTAGACAGTG ACTTCAGTC TAGGTTCGC TGCCGAATGG CTTCTTATTG GGGTAGTTGG ATTCCACAG TGCTGCCAC TGAACTGGT GTGGGGTGTG CCTGACACCTG CCCTGAGCTG ACTTCATCAT CTCTCTCCCA CTGCTCTCC AGCCTGTT AGTACGTTG TAACTCTTG GCCAGCTGT CTGTAACACT AACCTACCA TGTTATTGTT TAACTCTTG GCCAGTCTCT TCGTCTGAC CCTCTGAC CCTCTGAC ATGGACACT GACTGTGAT CCTGTCGCCA ATCTACCTCT GGACAAATTG CAGACAGCAA AAGGCAACTC TGGGGCTT GAGGACTGG CTTTGCAAA TTGGTTTC TGTTCCCTAC TGATATCTCA AGGAARACTCG TGGTGAAG TGTAACCCCT TTGGTACAC CAGTATGTC TCGAGAATGA AACCTACCA AGTCAAGGA TTGGAAAGA GATTATGTTT CGATGCCACC AAACCTCTG CAGACAGCC CACTTTTCT TGCGCTCTT CTGCTCTCTG CTCGCTCTG CTCGCTCTG CTCGCTCTG GCTCTGAGT TAAGGAAAG GCACTGGTTT AAGTTTAACT GACCCATCCA GSGCCCTTC ACTCTGTTAA CCACCTCTTC CTCCTCTG TSCTGCTTC AAGTGTCTCT GTGGCTGGAC TTTCATCAT TTGGGAAA CAGAGAGGC CTGAACCGG TGGCTCTACT CCTAGACCC CTGCTCTCT CTCATGGCTT TATAAACAGC TGTCTCTACT CAGTCTCTA TGCTCTCTA TGCTCTCTT CTGGGAGCA CTTGCTCAC TCCCTGCTAG CTGCTTAA AGGGCACTT CTGATAGTC CTGAACTCCA GCTCCCAGGC AGATGACTCC TTTA | P21462 | FME-T-LEU-PHE | 350 | 26 | 62 | 37 | receptor | |
| 189893 LG699 | Genomic Clone | 720 | AATGGGCACT TTGGGATG TGCTCTCTGG AGCTTAAGAA GAGGGCAC TTCTTCAAAAT GCTCTCTCA GCTCTCTCTG CTCAGGGTGT GGAGCTCTCG TTAGGTGCA CAGGGTTGG TGAGGATCA GCCCTCAGGG GCCTCTCTGG GGTGGGCTTT GTGAGGTTGG TGAGGATCA GCTCTCTCTG AGCTCTCTAT TTATCTGTTT GGTAGGCTCTG TCTCTCTCTG CAGGAGGTTG AGCTCTCTAT TTATCTGTTT ACCATGCTCTG AACGATACAG GCCTCTCTCTG TAACTGGCG TGTCTGATGC CCTGCTGCTT TTCTGAGCTC TTCTGCAACA GGTAGCTGAG GGGGGGTCC TGCCGGCTGG TGGGTCTAA AGCTATGAG AGCTCTCTT GTTGGGGTTT CAGTGGGTTT AAGGGATAGA GCACTTCAAA GGCCTATGAG GCCTCTCTCA GGCCTCTCTT GTACTGGCTG GIACTGGCTG TTTCTCCAC TCTCTGCTG CTAGCTTGA GACGACTCAC AACCTGGTTG CCCTTGCTCT CCTCTGCTG ATACAGAAC CACGACCCCT TCCTATCAC AGAAATCT AGTCTCTCTC AGGGCACCC AGGACAGGA AACGGTGTGTT TATAGGTGCG ACGCGGAGC AGCGGCA GAACAGGTT GTGGAGGA AGCTCCATT TGCCCTCTC ACTCTGCTCA CAGTATGAA AGTACATARA GTCCTGCTAC ACAACGGTTC CACTGCTCATC | 094910 | KIAA0821 protein | 1474 | 212 | 447 | 240 | 33 | |

| LS Cluster ID: | LS Cluster Name and Repre- | Sequence Length | From | To | Aligned Percent | | |
|--------------------|----------------------------|-----------------|--|---|-----------------|------------|-----------|
| Acc. No. | Homolog Name | Len. | From | To | Aligned Percent | | |
| (188894) LG1446 | Cluster Name and Repre- | 660 | | | | | |
| 190701 AL121834 | Sequence Length | 660 | GCTTGACATA GCTAAGCAG ATATGGCGA GTTAACTAAT TCCCGAGCA ATCAGGAGTC GGAADCAT GTCTGATCAT CTTGTTCTCT GTCTGGATGG CTGGCATTCT GTCTGAGATA CCCAGCTGG TTTTTTATAC AGTAATGAC AATGCTAGT GCAATCCCCT TTTCCTCCGC TACCTAGGA CTCAATGAA AGCATGATT CAATGTGAG AGATCTGCT AGTCTGCTGAT TGCTATTGTA TAAATATGG GETGIGETAC TTTATACGG CAAGGACAT CTGAAAGATG CCAAACATTA AAATATCTCG ACCCTTAAA GTCTGCTCA CGTCGCTAT AGTTTCAATT GTCTGCTAC TGCCTTATAA CATTGTCAAAG TCTCTGGAG CCATAGACAT CTCATCTCC CTGATCACA GCTGCAACT GTGCAAAGC ANGGACATG CTCATCAGT CACAGAAAGC ATTCAACTCT TCAACACTG CCTCACCCA ANCCTTATG GCTCTATGG AGCATCTTC AAAAACTAG TATGAAAGT GGCAGAGAA TATGGCTT CCTCTAGGTT CCTACAGGC CAACTGATC GTGAGGAGT TTCCCTTGA TTCTGAGGT CCTACAGGC CAACTGATC TTTATGATT | P49238 Probable G protein-coupled receptor GPR13 | 355 123 | 304 182 | 38 182 |

| LS Cluster ID | Accession No. | Database Type | Sequence Length | Sequence | LS Cluster Name and Gene Identifier | Homolog Name | Homolog Accession No. | Percent |
|---------------|---------------|---------------|-----------------|--|-------------------------------------|--------------------------------|--------------------------------|---------|
| LG Cluster ID | Accession No. | Database Type | Sequence Length | Sequence | LS Cluster Name and Gene Identifier | Homolog Name | Homolog Accession No. | Percent |
| LG745 | AC011780 | Genomic | 1245 | TGGCCTGC TTGGCTGG GTAGGGTGG GAAACCGAG GACTCAGAAG GFR61 | P50406 | 5-hydroxytryptamine 6 receptor | | 32 |
| | 189895 | Clone | | GACGCACT CCCGTAAG AACCCAGA AGTTCCTCTC AATGGAGGCC TCCCCTAG GCAGCTCAG CTCTCCCTGT CTGGCTGGT AGAACATCC SEQ ID NO: 22 GACAALCTGC TTGGCTGAGT CCCCGGAGT AGAACAGTG AGAACATCC SEQ ID NO: 22 AGAAAGAAGG GTTGGAAAGT AACGAAAGT AGAACATCC SEQ ID NO: 22 CTCTCCACCT GCCCGTTGA ATGGCTGTA GCCTCTAGGG CAACATAGAG GTGGAGAGG AAGTGGGCA ACCAACAGAG CAGAACACTGT CCCTCCACAG CAGAACAC CACTCTGCT TTCCCTCCC CAACAGTCCG GTGGGGGGTG GTCTGGGGGG CCCCGGAGCT GGAACATCC GGGAGCGC TGCTGAGAGA TTCGGGGCTG TGCCGGGGT TCAGGATCC CTCAGATCC CGCCGCTGCT GCATGCAAGC CACGGGGCC ACCAGGAGA TGCTGAGTA GACCACAGT ATGAGAGGCA CGGGGAGAC AACAGTACG AGCTGAGA AACACCAAAC AAGCTGGAG TAGGGCTGT GCTCTCTGT GAGTGAACAG CCTGGGGGA CACTGGAGC TCCCTCTCC CAGGAGACCC TTCCCAACAC TGCCACGAA GCCATGCCA AGGGCTCAC CCACACCC ACCAGGAGCT AGGCCACCAAG CCACACGTC ATGGCACTC CTGAGCTGT GGCTGGCAT AGCTAAATG AGCCTCCAC ATGGCTGT GACCTGGCT GACCCGAGA GGTGGCCAG GCTGACAAAG CACACCTCA GAAAGAGTA GAGGGGGAG GCGACCTCC CAAAGGGGC GTGGTAAAG AGGGAGAGC TTGGAGAGT GCGAGGGC ATGAGGCTCA GGGAGCCAG CAGTCACCC AGGGAGGT GAGAGCAA GAGLATTIT CGGAGGGCAG CGGTCTGGC GATAAGGGC ATTACAGGG CTTGGCCAGC CAGAGAGTC AAAGTCAGCA GGACATGAA GAGAGGGCC AGAGATTCG AAGCACATC CGGTGGCCCC ACCCTGGGAA CCTAACCTGC AGTAGGGGA CCTGGGGCTT GAGGACCTT CCTGAAAGTG GAGAGTCC CTGATGACTG GGGGATGGGT GAGGACTCA TGGGACTCA AGGGGCCAC CCGGG | | | | |
| LG1440 | AL121755 | Genomic | 477 | ICACCCACAT CCTCTACCTGT CAATGGCAT GGTATGGGG AGACGTAAG GGAGCGGTG CGCGGGTAGT TGCGGGAGC AGAGACAGC TGCCCTGTGT CCCAGAGAG CTGGCTGACCT ACCTAGTGT COATCTGAA SEQ ID NO: 23 GGGCGCCAG ATGAGGCC CAGGAAGTC GGRATGSC AGTTGCAA TGAGGGATT GTGGGGTGT GCGAACTCT TAAGGGGT GGGGCAGGG ATAAAGACAA AGTTCACAT CCCCCAGAC AGATGANGC CGGCACTGC AATGGCAAGT AGCAGCTGG CTGGGAAGAA GGCGGGCTC TTGGTATGT CCTCAAGTC ATCCATAGG AGGCTAAATC CAGCATACI GAGTTAAG GAGAGGGAGG AGGGATGGTC TTGGGGTCA TAAAGTGG GTGTGAAGCT CGTGTTCACCA TTCTGGCTG CCATGT | GPR73 | AAD00248 | Neuropeptide Y receptor type 2 | 34 |

| U.S. Cluster ID: | Acc. No. | Database Type | Sequence Length | Sequence | LS Cluster Name and Repertoire Sensitive | Homolog Acc. No. | Homolog Name | Len. | From | To | Aligned | Percent |
|------------------------------|----------|---------------|-----------------|--|--|---|--------------------------------|------|------|-----|---------|---------|
| 189897 (189898) LG1439 | AL121755 | Genomic Clone | 792 | TTCGTGACA CGAAGTGGTC ACTTCAAGGCC GATACAGGCC ACCCTTGCT CCCCTGGCNC CCCGTTGGT CTGAGTCAA GTCTAGGCTT GCACTTGCTC CATGGTGTG AGGGAGCCA GTGCGAGCAGC ATCACATCTCT TGAAGTACTCT TGCTCATGTC GATGCACTCG CACCGATAGA AGCGCACTGAG GAAAGTCTGTT TCCCTACGA ACACAGTGG ACCACGTTGA AGCGCACTGAG GAAAGTCTG GAAAGTCTGTA CGAACGAGTCA CGAACGATGG TOAAACCGTA GAAAGTCTGCA CAGCACAGCA CATAGGCC GAAAGTGCCTA ATAGACCCA GGACCGCTT CCTGGCGCAG CGAGAGCGCT TGCGGAATCTG CTCCGTCTGG AATCCAGGA CGCCGCTGAA CCAGAGCTCC CGGAGATTC CGGAGATTC CAGGGTCTG AGTCAACCAG GGCACCAAGAA CTGACACCA AGATGAGA CGAACTTACCA CTTCCTACTG AGCCCTCTGAT CCACAGGCCA GNTCTGGCCA CAGAACGCTT TCTCTCTGCT CTTGACAAATA AGAAGGACCG TTTCCTGTGC AAAGTAAAGCC GATGGGATGG CAATGAGAAT GGACACCATC CGAACGAGG CGATCAGGA GGAGGGCTT TGATAATTCA TCCGGGGTTT CGAGGGTGA ACCAATGGGA GATATCTGGT GGGGGAGGGA AGCCAAACGTT AGTAATGATG AAATACGTGAA AACGTTAGTT TGTAACAAAC CCGAAACATA CA | GIR73 | P25103 | Neuropeptide Y receptor type 2 | 407 | 130 | 317 | 193 | 23 |
| 3008 (189899) LG762 | AC011923 | Genomic Clone | 303 | GAAGGCCCCA ACCCCACATT GCGCTTGGT GCGCTGGAG GCGCTGGG ACACATAGT CCATCCAGT GCCCTCACGA TETTCACCAT GGCCTGGCC TGCTGGAGT CGGGGGCAC CACCGGGAG AGAAGTCAT AGCGTGTGA glutamate GTGGCTGAG TCCTGGGGCTG TGGAGGCTCA GCTGATCTG GGTATCTGG GGCGAGGAG GACAGCTGGT CTGGTAAGG AGTCAGCTAA CTCAGAGGAG GGAGGGTAAAG GGGGGCCAG GACCGGGAGS GTCACAGAA GTCUTSTGGC ATG | O15303 | Metabotropic Glutamate receptor 6 precursor | 877 | 157 | 233 | 77 | 89 | |

| LS Cluster ID: | Current (Original) LG NO. | Acc. No. | Database Type | Sequence Length: | Sequence | Cluster Name and Repres. | Homolog Name | Homolog Acc. No. | Percent | | |
|----------------|---------------------------|----------|---------------|------------------|--|--------------------------|--|------------------|---------|-----|-----|
| | | | | | | From | To | Aligned | | | |
| 189901 | LC895 | AC013396 | Genomic | 1218 | GGCCCTTACCC CCAACGGCT GCAGGCCCTGC AGCGGCGCCU CGGCCCTUGGG ATGGCCCTTC CTTTTCGAGA GAGCTTCGGC CCTTCGACTT CAGCTTCGCT TCCTGGAGAG CTCCTCGCG CCTGGCCACT TGGGGCTAGA GAGCTTGAG CCAGCAAGC CGGGCTCTGG TGGGGCTAGA GAGCTTGAG AGAACGGGT CCAGGAACT GTTGGAGTAG GTGAAGGCTCA GGAGGCCATG GAAGAGGCTG GTGGCAGGTT CCAAGGATGCG AAGAGTATGCG TGGGGCTAGA CCACCATGCGA AGCCATGCCA AAGAGTATGCG TGGGGCTAGA GAGCTTGAG TAGAGCGCCA CCACCATGCCA CAGGATACGCG ATGGCCCTCTG GGGGGCTGCG CTGGCCGCC AGACATGGT TCCGGATGGT GAGGCCATG CTACACATAG CAAGAGGAT GAGGCCACT GGGCCAGA ACTCCAGCA ACTCCAGCA TGCTGCCAGC CGAGGCCAGC CGGGCTACCC TGAGCTGAG TGAGCTGAG CGAGGAGGG CGGGAGAAGG TGGCTAGGAG CAGCTGGCG CGGGAGGCGA GGATGCCAC CGAGACTCC CGGGCCACCC GGAGGCTGCG CCACACGAA GCACGGCTCA GCACCTGGTG GGGCTGCAAC ACCCTCAAGT AGCGGTTCAG TGCGATGGCT GTGAAGAAGA CAACCTGGC CGGGCTGGT GGGGAGGCGA TGAAGAGGTT GACCTTGCGA GCACAGGCCA CGAGGCCCA GGTCTCTGG AGGAGGTAGT AGTCAGGG GAGGGCAGG TGCTGATCA GGAGGAGTC AGGGGCCACC AGGGTACCA CGAACACCGT GTGGAGGAGT CAGGCCCG TGTTGGATGCA GAAGATGAG AGGGCAAAAC TGTCTCCAC CGGGCCAGG ACAACCTCA CGGGCAGGGAT TGGTCCCAAG AGGGAGGAGA CGAGGAGCA AGAGCTGGG TGCGAGGGC CTCAGAGGAA CGGGCCACCA CGAGGCCAGG CHAGGGAGG ACAGGAGGGT GAGGGAGAA AGGGAGGAGG GAGAACAGAG GGGGAGGAG AGGGAGATGG AGAGCTCG TTGTAAGT CCATGGGCTG CTTGGGCGCAT CGGCGCTG | P49019 | Probable G protein-coupled receptor HM74 | 387 | 18 | 301 | 280 |
| | | | | | | | | | | | |

| LS Cluster ID: | LS Cluster Name and Pronic | Homologous Gene Name and Pronic | Homologous Gene Accession No. | Homologous Gene Sequence (SEQ ID NO.) | From To | Aligned Length | Percent Identity | | | |
|----------------|----------------------------|---------------------------------|-------------------------------|--|---------|----------------|------------------|------|-----|----|
| 190411 | AC011457 | Genomic clone | LG5580 | <p>ATATAGTTC CTTCTGAT GAAAGGAT CACATGAC TTGGGAGCM. Phenotypic receptor precursor</p> <p>AGATGAGCC AGTAGACCA GTCGTGACG CCAGCTGCC TGCGTCGTT AGTAACCTGG GAGGAGGCC SEQ ID NO:27</p> <p>GCGGCACCTCT GCCTTGCC ACCAGAACAC TGCGAGACG AGTCAGCTG AAGTGTAGA AAAGGCTACA GCAAGGGACC</p> <p>GTGGAAGCA TCAGGCACT TCCCTGCGCA AAGAGCTACA GCAAGGGACC</p> <p>CTAARACCCA TAGCGTCAA GAGCTGAGA CGAGTGAAGA GGCAGCTGCG</p> <p>GAGECCTCAT TAGACTGCTAT ATAGCTGAGA CGAGGCTGAGA ATCTGAGTC</p> <p>CTTGTAGG AACGGAGCT CTGCCCCG CGAGATTCA CAGAGGGTCA</p> <p>CTTGGTCAA GGCCMAGG AAGACAACAG AATTGTTT TGGGGAGCC</p> <p>ATCCACCTTC GGATCCTGCT ATAGCTGCTT TTCTGGCTT AAGCTGCTCT GTGAGGGCTA AGGGCTGCTT GGAGAGG</p> <p>GAAGACACG CCACACTCA CACAGCTCA AGGGCTGCTT GGAGAGG</p> <p>GCGCTGGCA AGGTGGGCC GCGCGTGAAGA GAGCGAAGC</p> <p>AGACGCTGAG GGAGACAGG AGGGTGTAGC TCTAGAGCTG GTGGTGGCC</p> <p>TGAGCTATGG GACTGTGTT GTCCACATA RACTCCACA GGATGAGAAC</p> <p>AGAGGAGCA AGGCTGTCA TGCCAGAGC AGGGCCAAAG</p> <p>TTTCATCATA GCGCGAGGA</p> | 1078 | 706 | 850 | 145 | 42 | |
| 190412 | AC010136 | Genomic clone | LG5459 | <p>CAGAAATCC CAGCTTCCAC AGAAATGAC ACGTTTCA AAATAAAGC</p> <p>AAGGCAAGCT GTCCTACCC AGAAGAAAT CTAGCTAGCA AAGTGGCTT</p> <p>CCTCTCTGAG GCCCGAGCCA CCTGTCGCA CCACAGCTCA</p> <p>GGAGTCAAGG TGTACTTACA GTTAGAGGCC ACTTCATGAG TAAGGTGAA</p> <p>TAAGGAAACG AAGTCAGAGC ACACCTCCCT TCAGAAGTCCC AACCATGTC</p> <p>ACATCTGGG AGAACAGTT AAGTCAGGG ATCAGACACI TGTGTTTACA</p> <p>GACTGCGAGG GICCATTAGA GTGAGCTGG GTCCCAAGCTG TGAGCTGG</p> <p>GTTGGAGGC CATTATCTGA ATTTCCTACT CTATGATGA TCACCTTTTNT</p> <p>TCTTTCTCTP TCTGTGATT TATTTCATC TTGTATTCG TAARTTCCT</p> <p>GGPAGATCAC CTTGTAAGAC TGGCAACTGT CTGTAAAGAA TAAGGGGA</p> <p>AGGATTTGAC TTACAGG AGACTTCA AGGACTCCCTC TCTAGGAGCA</p> <p>ATTTGGGGC AAATCCAGTGG GAAGGAGCTG GAAGACTGCA CTTGAGCTC</p> <p>GTGAGGAGCA CTTGCAACAA ATCTTCACT ACUTTCAGG CTGCTTTCAG</p> | 1.9592 | 9269 bp | 3917 | 3266 | 654 | 89 |

| LS Cluster ID | Acc No. | Database Type | Sequence Length | Sequence | Homolog Name | Homolog Accession No. | Sequence ID (SEQ ID NO.) | LS Cluster Name and Repetition | Percent Aligned | Percent | |
|------------------|----------|---------------|-----------------|--|--------------|--------------------------------------|--------------------------|--------------------------------|-----------------|---------|----|
| 190414 LG5853 | AC016468 | Genomic clone | 1575 | CTTTTTCAC CAGCATAGGC ACTGAGTCGG GTCGTGCAAC CCCTTGCGCA CCACCGGTT CGGGAACATGA GTGCCAACCTGC ACCAACACAA CCCGGCAGAG TAACACAGC TTCGCTTGAC TGCCACCTGC CAAATGCCC ATCAGCTGG CCACCGGCA CACAGCTGA TGCCCTCTC CATCGCTCA ACCCGTGGTG TTAATCTGCT CGCGCTCT ACATATGCTG GGCGTAGTG TGGAGCGCA AGCGCAAGCT GCCTGAGGTG ACCAACGCTT TTATCTTATA CCCTCTCTA ACCGACTCTGC TGCGATATTTC GCTCTGGCC CCCCTGGGTTG TGCACCTCTC ACCGACTCTGC TGCGATATTTC TCRAGGICCA CTTCGCGAGG TCAACACCTG TGGTGTGTT GCCTGAGGTG TTCGGCAEGG TCAACACCTG TGGTGTGTT GCCTGAGGTG CATCAACAC CCCTCTCTCT ACCGCTCAAC GATGACCGAG CGCCAGGTT ACCCACTCTCT CTATCTTATA CTATCTTATA CGATCTCTCT GCTCTGGCC CCCCTGGGTTG TGCACCTCTC ACCGACTCTGC TGCGATATTTC CTCCATGATC TGGGAGCA GCCCACTGTA CTCATCTTC ATGCTGCTG CCCTCTCTCT CTATCTTATA CTATCTTATA CGATCTCTCT TTCCTGGAG CCCGGAGGA GATGACCGAG CGATGCTCT CAGCTGGAA CTGGAGGAGCA ACCGCTCAAC GATGACCGAG GAGGAGAGA GAAGGAGGAG TTCAGGAGT AGATGAGTT TGCCGCCAG CATGAGGCTG AGGTAAGGC CAGGGAGGC AGATGAGAAG CCAGGAGG CAGCCGGAAG GCAAGGAGA GAGGAGGAG GAGGAGG AGGCCGGGG CACGGAGGA CTCAAGAGA GCAAGGAGG GGCAAGCTGG AGGTAAGGA AGGCAAGCC AGGCAAGCC GAAGGAGAG AAGGCTGGCA CAGGGCTCA CGATGAGC ATGATGTTG GTGAAGATGA CAGGGAGTT GGTGAAGAGC AGATCAATT CAGTGGAGGAT GACGTGGG CAGGGAGAT CCGGAGAGC CTCGAGGCA GTGCGTAA CAGCAACAGC AACCTCTCTC TGCCAGCTG CTACCAAGTCG AAAGCTGCTA AAAGTATCTT CTCATCACTT TCTCTTATG TGCTATCTCT TGCCTTTAG CAGCTGGC CAGCTGGCTG GATGCGAAA CCAGCTGACC CCAGCTGGTGTG ATCACCAATA TCATCTCTG TCTCTCTG CAGTGGAGCA TCCACCCCTA TCTCTATGSC TACATGCACTA AGACCAATA GAAGGAGATC CAGGAGCTGC TGAAAGAGT CTCGGAGA GAAAGCCC CGAAAGAGA TAGCCCCCCA GACCGCCCG GAAAGAGA | P08912 | Muscarinic acetylcholine receptor M5 | 532 | 2 | 518 | 470 | 23 |
| | | | | | | | | | | | |

| DS Cluster ID | DS Cluster Name | DS Cluster Length | Sequence | Homolog Name | Homolog Accession No. | Sequence (S+Q, ID No.) | From | To | Aligned | Percent |
|-----------------------------|-----------------|-------------------|--|--------------|-----------------------|------------------------|------|----|---------|---------|
| Acc. No. | Type | | | | | | | | | |
| 199886 (90416) LG6804 | Ali36961 | 1001 | TGGGTTAATG CCACGCTTCA AGCCAGTG GCTGAAGGG ACAGTAAAGA AGCCCTCCGC TCGGCACAGG ATGGCAGTGT GAGCATTCCT CTTCGCAATGA ACTGTTGAT GTGGAGTGTG TAGGGGTGA AGCACACC CACGGGACCC ACCAACAGCA GCTTAEGAG CGAGCCCTCGC CAGTSGCGTC CTTTCCTGCTG CTGTCACCTGG TGCTCCCGC TCCTCCGTC TTGTCCTGCG TTTCCCGCTG GTCACTGGT CCTCTGGT CTCCTGGCT CTCCTGGCT TTCCTGGTGG TCACTGGTC CTCCTGGCT GCCTCCCGG TGGCTCCCT TCCTCTTGTG CTCCTGGTC TCCCTGGCT CCCCGGGGT GTGTCCTT CCCGGGGGTC CTCGGTCTT CCAAGCTG CTGACAGCT TCACAGTGAT CTTCATAATGC CAGGACAGGA TATTCCTAAC TGGCCCCACAG AAGCCAAATGG CMAAGGCCAC CAGGACACAG AGGGGACCC GAGGAACTGA CTCCATGCG CTGTAATCCA TCACTGGCTT CTCGGCTTGG AGCGCTTGGT TCACTGGCT CAAGGACAG GGCATCTGGT GAGCAGCAC CAAGCTCAG ATGGCAACCC AGACGGACCT GGGCGGGCA GCGGTGGGA GGCCTGGGC CCACTGGGA CAGACACAG CTGGTAAATG GTCACGTC ACACAGCCA TGAAGTAGAC CCCCCTTGG TCTCGTCTTA GAGCACAAAC GGCCTCAAGCC TGGAGAGCC CTTGCGCAA GGCAGCTG ACCCAGAC ATAACACACC ACCCTTCCCG GTAAGGCCAC GGTGAACAGC AGGTGAGACA CAGGAGGTG CACCAAGGG ATGGCTGTGC AGPTGATCTT CCTGCTTGA TGAGCTGA GGTAAAGGGC AAGGATGTT CCCAGGGAC TGAAGGACG GAGGCTGTG TAAACAGAG | P2249 | 361 | 36 | 118 | 83 | 44 | |
| | | | | | | | | | | |

| LS Cluster ID: | Accession No. | Database Type | Sequence Length | Sequence | S. Cluster Name and Repetitive Sequence (SEQ ID NO.) | Homolog Accession No. | Homolog Name | Len. | From To | Aligned Percent | | | | | |
|------------------------------|---------------|---------------|-----------------|--|--|--|--------------|------|---------|-----------------|----|--|--|--|--|
| 139889 (109417) LG5881 | AC016856 | Genomic clone | 947 | TGCTTGACCA CGGAAGAAATT TCCRAAAGGT TCCCACACTCT GTTGTCTGAC CCAGTTCTTC TTTCCTCCCT CAGATAGTGT TCGATGTTAG AAGTGTCTG ACAACCAAGTG GCCCCATGTC CAGAGGGCG ACTGCAATCC CAAAACCTT SEQ ID GACTTGTGTG TCTAACACA GCGCTGTC ACAGCGTGC CTGTCTGAC NO:21 AGCCCGTC TTTCTCCTAC TCTAACACA GCGCTGTC ACAGCGTGC CTGTCTGAC ACCACACAC ACCACACATC CGAACCAACAA ACTGCCACCT CGAACCAAC CTGGCTTCCTC CCGTGGCCCT CCCATGAC CCGTGGCTC TGGCCCTCA TTTCATGG CCACCGAGAC CCCATGAC CCGTGGCTC TGGCCCTCA TTTCATGG CTCTTGTTGTG CTGACCATCC ACTGCTGTC CGAACCAACAA ACTGCCACCT GCAGCTTCC ATGCAACCC GCGAGACCC CAGCTTASGG AGCTGGGGGG GRAGAGCTC CTGACACCA TCCMCACAGT TCCCTGACCC AGGAGCCCT GTGGAATC TGGGTACCA GATGGCCCCC TGGCCCTGT AAATCTACA GAACCTGGC CCACAGTGC TSPAAAGTGT GATAAAGGCT CCTTGGAACT TCTCTTGAA CTGGCTTACT TGACTTGTGT AGATCTGTC AGCTTGCTGG TGACCTTCCC CACCTGCGG CTGCTGACCA CCTCTCAANG AGCAAAANGCAT ATCACACATCA GCATGTGTC TGCTCCTGT TGCTCCTGT TGCTCCTGT CCCTCTGGCA TCAGAAGGAG GCTCATCTC CTCCTCTGT TGCTCCTGT CATCTCTCAT CCTCTCTCAT CCTCTCTCAT CCTCTCTGT TGCTCCTGT GCTACTCAT CCTCTCTCAT CCTCTCTCAT CCTCTCTGT TGCTCCTGT TGTTAAATG GAGCCTATG CTGGCTGGA TTGGCTGCTC ATGGCTCAT AGAGCAAGG GTTGTGCAA CCATGGCTC AGGTGAGGT GCGAGCAGC GPR 84 ATGTGGACCA CGCGGGAGC CTGGCTACTG AGCTCCAGAA GCGAAAGACA SEQ ID CAAGAAGGG ATGTGACTCA GGAAAGGCA GAGAAACACA GCGAAAGACA EX33 TTGAGTCACTC CTTCCTCAAT TCCATGAG ATTCGGGGC TTTCCTGCG CCCTTATATG GCTTSCTCTT GEGGATGTT TGCTGAGGGC TTTCCTGCG CCTCTCTTA CCTCTCTTAC TGTGATGTT GTCCTCTGT GTCCTCTGT CCCCCTCCAG GGTCTGGCTG GTGCTGCTC TGCTGCTC AGATGAATC CCCTCTGG GTCCCTCTGA TGCTAACCTG CTGCTCTGT CCTGGAAAG ACCAGGCTG GCCTCATCTAG TCCCTGGCCN ATGGTGGAG TGATGCTTG CCTGTCGAA CTGCTATGG TCCCTGCTC GTCCTGCTC TTGTGCTGG CGGTGATGA GGCAATGAA GATCCAACA CTCCTGACCC CAGACACAA GTAGATGCCC ATGAGGATGG TGCTGTAAG CGGGCTCTGG ATGGCTGAA AGCTGGAGGT CGAGACTACA GTTACAGGA TATAAATAGG CGAGGGGA GCAAGCTGG CCACGCCAC AACCCAGTC CGACCCAGT CGACACTAT CCCTCTGGCA CTGAAACAT GGGAAGAAG CTAGGCTGG CGATGAGGA GCTAGGCTCC CAGTGGCTG AGGGAGGGC TGCGGCTGA | P41180 | Extracellular calcium-sensing receptor precursor | 1078 | 577 | 828 | 246 | 36 | | | | |
| 139418 LG6808 | AC020641 | Genomic clone | 840 | TITTAATG GAGCCTATG CTGGCTGGA TTGGCTGCTC ATGGCTCAT AGAGCAAGG GTTGTGCAA CCATGGCTC AGGTGAGGT GCGAGCAGC GPR 84 ATGTGGACCA CGCGGGAGC CTGGCTACTG AGCTCCAGAA GCGAAAGACA SEQ ID CAAGAAGGG ATGTGACTCA GGAAAGGCA GAGAAACACA GCGAAAGACA EX33 TTGAGTCACTC CTTCCTCAAT TCCATGAG ATTCGGGGC TTTCCTGCG CCCTTATATG GCTTSCTCTT GEGGATGTT TGCTGAGGGC TTTCCTGCG CCTCTCTTA CCTCTCTTAC TGTGATGTT GTCCTCTGT GTCCTCTGT CCCCCTCCAG GGTCTGGCTG GTGCTGCTC TGCTGCTC AGATGAATC CCCTCTGG GTCCCTCTGA TGCTAACCTG CTGCTCTGT CCTGGAAAG ACCAGGCTG GCCTCATCTAG TCCCTGGCCN ATGGTGGAG TGATGCTTG CCTGTCGAA CTGCTATGG TCCCTGCTC GTCCTGCTC TTGTGCTGG CGGTGATGA GGCAATGAA GATCCAACA CTCCTGACCC CAGACACAA GTAGATGCCC ATGAGGATGG TGCTGTAAG CGGGCTCTGG ATGGCTGAA AGCTGGAGGT CGAGACTACA GTTACAGGA TATAAATAGG CGAGGGGA GCAAGCTGG CCACGCCAC AACCCAGTC CGACCCAGT CGACACTAT CCCTCTGGCA CTGAAACAT GGGAAGAAG CTAGGCTGG CGATGAGGA GCTAGGCTCC CAGTGGCTG AGGGAGGGC TGCGGCTGA | P23566 | 5-hydroxytryptamine 1E receptor | 365 | 109 | 360 | 280 | 25 | | | | |

| LS Cluster ID | Access No. | Current (Original) LG NO. | Database Type | Sequence Length | Sequence | Homolog Name | Homolog Acc No. | From | To | Aligned | Percent |
|---------------|------------|---------------------------|----------------|-----------------|--|--------------|--------------------------|------|----|---------|---------|
| 190419 | AC021089 | LG6171 | Genomic clinic | 867 | GTTTGCTTCA GAGCTTAACC AGCTTTTCCT CTTGCGCAG CAATATCTT GAGAGTCAT ATCCGTCTCC AGCAGAGAC AGCTGGTGGC AGAGAGATCTTC ACAACPATCH CTGGCACACT CTGGCTGCTC ACTCTCTGG CCTCTTTTC ATAGTGTGTTG TGGACTTCCT GTGGAAAGT TGGAGAAGT ACATGGAGAT GCCTCAGGTC CCCCACAAAGA TCTAGAAGT GCCTGAATT TGGAAATTC ACACCTCCAT ATGGAACTACT GTAACTGTTAA CCTTGTAGAG GATAATGCT GTCTGGCACC CGCTCAGTAA CCTAACGGTC TGTACCCAG CCCTGACCC GAAAGTCATT GTAACTGTTT ACATCACCTG CTTCCTGACCC AGCATCCCT ATTACGGTG GCCCAACATC TGGACTGAAG ACTACATGAG AACCTCTG CTTCACCTCC TCATCTGGAT CCTCTGGTC ACCTGTCTTC AGCTCTTAC CCTCCATCTTC TTTCATCTGA ACTCATCAT TGCTCTAAG CCTCAGGAGA AGAGGAAATT TGGTCCTGGT GGCTACCTCA CGGGAGAGC CACGCCATC TGTGTCACCA TTACCCCTCAT CTGGCCACCA CTGGGGGGC CCGGGATCAT CATGATCTT TACCACTCT ATGGGGGCC CATTAGAAC CCGTGGGG TACACATCAT GTCCGACATT GCACAACTGC TAACTCTCT GCACACAGCC ATCAACTCTT TCCTCTACTG CCTCATCAGC AACGGTTCC GCACATGGC AGCCGACGCG CTCAGAGCTT TCCTCTAAGTS CCTGAGAGCA CCTGAGCT TCTACACCA TCATAC | 014694 | CCDS receptor (fragment) | 333 | 22 | 304 | 270 |
| | | | | | | SEQ ID NO:30 | | | | | |

| LS Cluster ID: | Accession No. | Database Type | Sequence Length | Sequence | Cluster Name and Reference | Homolog Acc. No. | Homolog Sequence (SEQ ID NO.) | Len. | From To | Aligned | Percent |
|----------------|---------------|---------------|-----------------|---|----------------------------|------------------|-------------------------------|------|---------|---------|---------|
| 190420 | AC021773 | Genomic clone | 1063 | CCCTTCCCAGG GAAAGCGAAG CGGGGCCGT CGTGAGCTC TGCCCTCTGA GGAGGCCTCC CGGAGATGG AGAAGTGG CATAATACA TCAAGGGAC AAGGTCCTCG CCAGTTCTCA AGAAAGTACA AGCAAGTCGA CCTCTCCCTG GCCTACAGTA TCATCTTAT CCTAGGCTG CCACAAATAG GCACGTGCTT GTTGGACTTC TGGGGCAA CCAAGGGCTG GAGGTGTGCC ACCACCPATC TGTGAACT GTGGTGCC GACCTCTT ATGTCATT GCCTTCTTC ATCATTCCTC ACTCATAGA TAGCATAGA 'GAGCAGTGG CCTTTCGGGG AGCTGCTTG CAAGCTGTG CACTTCCTGT TCTATPATCAA CCTTTACGGC AGCTTCCNC TGCTGACCTG CTTCTCTGT CACCACTTC TAGGTGAGG CACCCACTG TETTCCTGTC CCTACCGAC CGCGAGGCAAT GCGTGCCTG GCACCAAGNC CAGCTGGCC CTGGTGGTCC TCGAGTGTG GCCAACACTG GCCTCTCCC AGACGGACTA CATCAATGG CAGAACTATG GGTAGACAT GACCAAGCNA GAGAATTGG ATTGGTTT TGCTCTGGC ATTAGTTGA CATTGTTGA CATTGTTG CTTCCTTCC CTCCTTGTC ATTTCCTGT GCTTACCT GATGTCAG AGCCTGATCA AGCCAGAGGA GAACTCTAG AGACAGGCA ACACAGCCS AGCCAGETCC ATCCGGACCA TCCATCTGTG GTGTCCTCTT TTACACCTCT GTTGGTCC CTTCATGATC ATCGCTCTC TCTACCTAC CACCTGCTT CTTCCTTCC AGGACCTCCA GCTCTGTG GCAAGCCCTAG GTGCCCTACA AGATATGG AGCTCTGGT AGTGAGCA CTGGCTCAA CCATCTCTG TACTTCTTT CAAGGGGGC AAAATAGAG TCAAGCTCTT CCAGAAACTG AGGAGAACTCA AGTTGGTGA GCACTCCAGT GGAGGAGAAGA GATGCCAGG GTTGAAAGA TCT | P2Y purinoreceptor | 328 | 4 | 202 | 93 | 43 | |
| LG629 | | | | | | Q15077 | 6 | | | | |
| 190421 | AC023078 | Genomic clone | 729 | AAGNNNNCC CAGCTGGTGNAT CCATTAAT AGGGAAAAAC TGATGCCA AGGGAGGCG AGCAGGAGG AAGACAGTA CTGAGGAGG GATGGTCAAG TACAGCTCTG TCGAGGGTAT CTTCAGGAGG CCACAGGAGG TCTCTATCG CAGGACAGG CNGACAGC AGAAAGCAC AATAGAAA NTCAAGCAG CTGACTGAT GAATCTGAT GTTTCACACC AAAGCAAAAC AGCACCACTG AAGAGGAAG CACATTAAT COACTCAGG ATGCTCGCA GAGGGAGAG GGGCCAGAGC AGGACACACA CCACCGCTGA CAGGTGTG GGGGGGTGGS AGGGTAACCA GATGGGGCA AGGACGAGA GGTGGTCAAG GGTGGTCAAG GCACTGAGAA AGCTCAAGCC TGCAGATGAG GAAACATCA TCAAGGATA GAGGATTTA GAGATGGTAT GGGGAACT GATGAGCT AACAGGAAAT ATAATAGCGC GCGGCTGAGG AAGGGAGT CTGCGCGSC CAGATGAGG ATGTTAGCTG AGAAGGCTT CCGCGGCTT CGGAGGAGG GGGCGAGG CAGAAGTSGC TTTCCTCTCA GCCCCAGAG GGAAGAGTGA CACCTCAGCA CCGTGAGGT CAAAGCTGTC TTGAGAAA GAGTCCTCTC AGTTCGTTG ATGGTGTCA GPTCTCTCA CAAAGCTGTC | P35410 | 378 | 70 | 282 | 210 | 38 | |
| LG6465 | | | | | | | | | | | |

| IS Cluster ID | Acc. No. | Current (Original) L.G. No. | Database Type | Sequence Length | Sequence | LS Cluster Name and Reference | Homolog Name | Homolog Acc. No. | Homolog RefSeq ID (No.) | To Ttron. | To Tlen. | Percent Aligned | Percent |
|---------------|----------|-----------------------------|---------------|-----------------|---|----------------------------------|----------------------------------|------------------|-------------------------|-----------|----------|-----------------|---------|
| 190427 | AL137118 | | Genomic clone | 1026 | TCCCTTTC ACCAACAC AACAGGAA TAACAGGA AACAGCTT GTAGCTGACT TTAGCTGTC CTTAAATTC TCGTGGATG GCGTTCTG AGTCAGACT TTAGCTGTC CAGGGATTG AAGCTGGCAT TGCGTGCCTC CAAGGCCAA AGTAAATAGAG CAGGGATTG AAGCTGGCAT TGCGTGCCTC CCACCTTCA TGCTCTGAG TGGCGGTTC TGAGCTGTC ATAGGGCAGS AAACAGAAGA AGAAGATGAT CAGGGATGAT ATGATGGTGG TGAGTGCTT CCCTGTGAGA ACCCGCAGCC CGGATTCG GACCTCCACT TTAAACAGA SEQ ID NO:31 CCCGTAAATGAT CGCGAATTA CAGATGCTGA GTGTGAAAAA TGCGGAGCG CAGCCACCA CCAAGGCAAT ATAGCTGCA TGTCGACCT TAGCATTTT ATAGAGTTAC AGCTCTAAC ATGATGTCAC ACTGCGCTC TGCTGAGC CACTGTGAG CAGGATATT GAGGAGCCA TGAGGAGGAT CCATATGATC CCACAGAGA TCCAGGACT CCTGAGCTG GTGAGATGCA GAGCGAGA GSSGTGAAAC ATTGCGAGA AACGACAAAC ACTCGCAGC GTCGAGAAT AAATACCT GTACATTTG ACATAGAGG ATATAGACAT ATCTCTGAG GCCAGGTCTC CAATATCCA ATTGGACCT CTAGATAAT AGTCAGCCCT GAAGGGAGC GTGCTATGA ACAGGAGAT TGAAATGCC AGAFTTAGCA TGAAACCT CAGAGAGT GACTCTAT AAGGTGCGAG GAAACATAT ATGGACACC CATTTCGCA GACTCCCCAG AAATATTTA TCAGATATAC AATGGAAA AATTCTCTC TGAAGTTTC AATTTGCGAG TTCTCTGCTG TGTTATTCGT GAGGGCGCA TTGGTCCCA TTTCGATAC GGAGATGGAT GGTGCAAGG ACATTAATT TCTCTC | Cystein-γ-leukotriene receptor 1 | Q0271 | 337 | 17 | 311 | 291 | 39 | |
| LG6807 | | | | | | | | | | | | | |
| 190428 | AP000440 | | Genomic clone | 426 | AAGTCAGCT GCGCTATTA TGAGCTGTTTG TGCTCTGAA GTCTCTCTT GATTTTATAA ACTGCTGAA TTTCCTGTC TGTTATGAT GCATCTGCT CTACTCTTT ATTAGCCCA TCAACATTT TTACCTGTC CTCTCTAGGC SEQ ID NO:32 ACTTTCTGCT CAGTGTAAAC ATCCTCTCA TTACATTAT CATCATGATC ACCTGTGATC AGAACATT TTGTACTC ACCATGGAC NATGATGAA CAACTGAGC CTCATATTA CTGTTAAA CATCTGTAT ATCCCTCTCT TCCAGCTTAC TGACAGTCTC TGGAGTATA TTFTTGTAT ATGTTAGGAG GTCAGACATT ATTTTTCTC CAGTGTACAT ACTGATCCT TCAAAGTTAC CACTGTGTC ATCATATTA CTGAC | U45983 | CCRS chemokine receptor (CMKBR8) | 1944 bp | 1941 | 1586 | 362 | 83 | |
| LG6894 | | | | | | | | | | | | | |

| LS Cluster ID: | Acc. No. | Database Type: | Sequence Length: | Sequence: | Cluster Name and Repres. Sequence: | Homolog Acc. No. (SEQ. ID NO.) | From To: | Aligned Percent: |
|------------------------------|----------|----------------|------------------|--|--|--------------------------------|-----------|------------------|
| 190430 LG5259 | AC003833 | Genomic clone | 549 | ATCATGAG AGTAAATT AAAAGACTTT CTCGAACAG AGCGAACAG CAGTGAAAG AGGCAAGGGT GTGAGAGGT TTCAAAGGA GTTAAAGATG ANGGATGCCA TATATGCTGT TGCCATACT TGAAATACCTT GGAACACAGT GACTAAAGCA ACAGTTGAGC ACGGCTGGG CAAGCTCTGG CCTGGRACAT CATTCACTGA TGATGTGAA CAAGCTGGT ACCTTGAGG ATTCTGPTATG TCAGTGAGG AAAAACTGAT GTCTAACCTC CTATCATATG CAAAAACAT ACCTTGAGG TCCATCAGTA AGCCGAAAGA ATGGATATC AAAGAGTTI TAAACATTA TAATATGTTT ACAGTTGAC CGATGTGAA ATAGCAAAAG TGATTTTAA TCAAGTGT TGATGAA TGACGTTTT AACACTCCGG TAAACTGCC TAAACTGCC TAACTGAGG GCTTATGAA GGACTAGGC AGCACTTGT CATAAGCA | U45983 CCR8 chemokine receptor (CMKBR8) | 1944 bp | 1328 1864 | 553 79 |
| 190419 (190431) LG5386 | AC008785 | Genomic clone | 894 | GTTATGAGT GCGGCGACT GTCAGGACT CTCCTGGAC TTGAAACAG CCCTGAGCT GTGGCTGCC ATGGCTGCC AGCGCTTCA AGCGCTTCA TGTTGCRAT TAGAGAGAA ACTTGTATGC TGTTTCAAGA AGCGCTTCA AGCGCTTCA GTGGGACATG ATGTTTACCA GCCGCGGTT CTCGGGGGC QCCTCCATAGA GGGGTAAAG ATTCATGATG ATGGGGGGG CCCTAAACTGT GGCAALAGATG GGGGTAAAG TGAACTGAGT GGCGGTGTC TCCGAGCTT GTACACATG ATGGATCTCA GAGAGGAA GATGGAGGAG GGCCTCCAGT AGACGGTGAAGA QCACTGGATC CAGATGAGAA CTTGAGGAGA CTTGAGGAGA AGAGTGTGCA AGAGTGTGCA GTGGGGCAC CAGTAAATGG GGATGCTGT CGATGCTAA CAGTGTGCA CAGTGTGCA GACTTCCC GTGGGGGGC GOTATGAGC CGTGTGTC GAGAGGAA TTGCTCTTC TCCGAGCTT GTACACATG ATGGATCTCA AGATGAGAA GATGGAGGAGA GGCCTCCAGT AGACGGTGAAGA QCACTGGATC CAGATGAGAA CTTGAGGAGA CTTGAGGAGA AGAGTGTGCA AGAGTGTGCA GTGGGGCAC CAGTAAATGG GGATGCTGT CGATGCTAA CAGTGTGCA CAGTGTGCA GACTTCCC GTGGGGGGC GOTATGAGC CGTGTGTC TTGAGGAGT GGCAACAGC GATGTTACCTG TCAATGTTTA AGCTTACAGT AATCCCTTTG GAGGTGAGA TGAATGAGAA TTCAGGACT TCTATGATCT TTCGGGAC CTTGAGGAC TCCATGTTCA TCCATGTTCA AGATGAAATC TTCCCAACAG AAGTCACAA ACATATGAA AAGAGGAC AAGAGTTCGG CAGCAGCAG TGCCAGAGA TAGTGTAGG AGCAACTTG 'TCCTCTTGCC ACACAGCTGGS AGAGTGTGAT CACTGTGAGS ATTTTGCTGAGA TGGAGAAGA AAAAACTGTT TAGCTCTGAA GCNAGATCT TTGTCCTCT ATAGGCGTAA GGTC | 014708 CCR5 receptor (fragment) | 352 | 24 323 | 286 26 |

| LS Cluster ID | Current (Original) LG NO. | Ac. No. | Database Type | Sequence Length | Sequence | Homolog Name | Homolog Acc. No. | From | To | Aligned Percent | | |
|------------------------------|---------------------------|---------|---------------|-----------------|---|--------------|----------------------------|------|----|-----------------|-----|----|
| 190705 (100432) LG5394 | AC008971 | Genomic | clone | 816 | GATTTTTT GGAAATAAA TATTTGCTTA GAAAGGGGT ACCCCAGCAG CTTCCTCTGC GAGTGTTAGA GGCAGCGCA GAACTGCTG TGCGGGCCA SALPR GCAACTTGTG CGGGAAACCG ACCAGTGGAC ACCTCTTCGC CCATCACCTT GACCGTGGTG GAGAAATGG CAGAAATGG CACTGGGCG CAGGCCAAG SEQ ID CCCAGATCCA CACAAACAGC GCCTTGGCG AGAAGCGCA GCTGTCCCCC NO:35 AGGCHICGGC CGCAGCAGTC GCCCTGGCG AGAAGCGCA GCTGTCCCCC CTTCAGAGCC GAGCCACCG AATGGTAGCG CCTCACACTC ATGGCAGTGA GGAAAGAACAC GCTGCGTAC ATGTCATGG AGTCACCAT GGACAGATC TTACACATGG CCTTGCCGAA GGCGCAATTG AAGTCAGAG CGTTCCTCCAC GCCCGAAGG GGAGGGTGA GAAUAAACTG AAAGTCGTC AGGCCAGGT TGGTGAGAA GAGGTGATA GAGGACTTC GCAGGCGCTG CATGCTCTTC ATCAGTGAGA GAGCTGGG GTTCGGCG AACCCAGG CGACACAC CCAGTACACC ACGTGTTAGA GATTCGGAC CGGGCTCTT CTGTCGGGC TCTCTCCCC CCCGCGTGGCC GGGGAGATG CTGGCGGCGC GCGTCCGGC AACTCGAGCC CCAGCTCCCCA CCAGAAGTTC GGAAAGCTCA CGGACCGGTT ACCACATGG TTGGCGGCTT CGAGAAGGTG CGGAGGCGA CTGAAAGTT CTGCPAGCTT GTCCCC | P41143 | Delta-type opioid receptor | 372 | 6 | 203 | 215 | 33 |

| LS Cluster ID: | Database Acc. No. | Sequence Length | Sequence | Homolog Name | Homolog Access No. | Sequence (S>Q ID No.) | From To | Aligned | Percent |
|----------------|-------------------|-----------------|---|--------------|--------------------|-----------------------|---------|---------|---------|
| LG355 | AC008754 | 1068 | CGAGTTCCT CCTCTAGTT TTCACTGCTT ATTCTGTCTAG CTAGAGATG GGGATTACA CGGACCTTC GGACCCCTC ATGGACCCC GETGGGGGT CCGGGGATG CCCGGCGAG GTGGGTGCC ACCCTGGTC TGGTGGCTT GCTGGCTCT GCGCTCTCT GCCTCTCTG AGCCCTGG CCGTAACTG CAGTGGATG CAGTGGATC AGCTGGTC CTCCTGCTAC GAGCTGGTCT TCTGACTCT AGCTGGTC CTCCTGCTAC GAGCTGGGG GTGCACTGG CCTGCTGGGC AGCTGGACA CGTGGCTTC TGCTCACGGT GCGCTGGCC ATCTACGCC GGCTGACCA GGAGCATTC CGAGCCGGC TGCACTGCT GTGCACTAC GGCTCTCTC CGACCTGG ACTGCACTC GGTTCTT TTGGTCTCTG TGCCCTGGTC AGCTGGCTTC AGTGGCTTC TGCTGGGGC AGGCGAACCC TGCCGGGGC TGGCAAGC CATTGGTGA GGTGTTTG TCCTGCTGGC AGCTTACAC TGCTGGGGC TGGTGGTCA AGTGGCTAC TGTGGGGCC CCAACTCGG AACTTACAC TGCTGGGGC CAGGGCTTG CGGGCTGAC CCCATCTGT GGCCCTGGC CTCGGCTCA CGTGGCTCA TCCCTGCTC TTGGGAGGGC TCAACTGGC CGTCACTGC CAGCTGGCC CGACGAAGT GTGGACAGCA AGAATCCAC CGGTTCTGG AGATGGAG | P21730 | 350 | 10 | 339 | 319 | 40 |

| IS Cluster ID | Current (Original) LC No. | Accession No. | Database Type | Sequence Length | Sequence | LS Cluster Name and Repetitive Consensus Sequence (SDQ-ID NO.) | Homolog Acc. No. | Homolog Name | From | To | Aligned Percent | | |
|--------------------------|---------------------------|---------------|---------------|-----------------|--|--|-----------------------------|-----------------------|------|-----|-----------------|-----|----|
| 190774 (190488) LG263 | ACU07922 | 540 | Genomic clone | 540 | CTTTTATA CAAAATTTC TCAAGAACG CTTTGAAAG CCTGATGAGGA ACATGGATA CAAAGAGGA TTGACAAAGG ATTAACCCA CTGANAGCC AAATGGATT CTATACCAA CTGATTTAGG ACCCTTGCT GAGGGATATA ATGAAGGAC AATTGGAAC AGAGGAAATAT GGAGGCCAGC AAACAGCATA IAACCCCTTA GAGAATGCC AGTGACTTGG CTRATCTCCT GSCTCTAAGC AGTTCAACAT GTTCCCCTTG GTGAAGAGCT ACAGATCTG ATGGAGAGA GGAAACCCATT TTGGAGCAA TTGTATGCT ATTGATCTG GTCTCTGAGG AAACATGAG ACTACCTTT CTCCTCTGC TCCTCGATG AAAGGATGCA GGAACTCTG TCGATCGAGA ANGAGATCTC CTGCTAACCTC GAATGAGGT CCACAGATGT TGAAGAGAC AGCACTAGT CCAGGATGGC TTGGCACT ACTGAGATGA TCACGCTTCC ACAGSCTCCA | Histamine H3 receptor H4 receptor (SEQ ID NO:37) | Q975NI | Histamine H3 receptor | 445 | 259 | 431 | 180 | 29 |
| 190557 | AI806860 | 574 | Dbest | 574 | TTCACAGAT TTATTGAT ATTGAGAT TATTAATGT ATGACATTC TGTTGCTGAT TTTCGAGAT ATTGGCTCTC CTTCATCTG AAGTGAGGAG CACAGAGGA CTGATCAGG CATAGACATT GACTACAAGA CTGAGACTT GCTGGTGTAT TGGGCATAT GCCCACAACTA TTGCTGAGGA GAGGGAGT ATGAAATCCA CCCAACAGAG AACCTTACCA GCAGGAGTAT GGCTGTGATG GCGCGTCTCTC CAGGGAGAG TCCCTGGAA GGACCATTC TGTTGAGGTCG GTGGGATGAC CTCCTAGCC TGAAATAGAG AGTCAACCCTG TATGCAATTAA AGAACAGCG TATTCCTAAC CCTPAAGT TGTCAGATA AGAAACCTGG CCCTGGAGAT GAAGCTCATG GAGAAAACCTG AGCAGTACCT AGTACATTTC AGTACATTTC AGTACATTTC ACAGGAAAGCA GCTAACGCTG AGAAGATCTC GTTACDACTG AAGAGCAAGT TGAGGAAACCA TATGAGAG AAACATGAT AATG | Q62851 | Pheromone receptor VN7, rat | 273 | 99 | 269 | 175 | 36 | |

WHAT IS CLAIMED IS:

1 1. An isolated polypeptide encoded by a nucleic acid molecule
2 comprising a nucleotide sequence that is at least about 80% identical to the sequence set
3 forth in Table 1.

1 2. The isolated polypeptide of claim 1, wherein the nucleotide
2 sequence is set forth in Table 1.

1 3. An isolated nucleic acid molecule, or its complement, encoding the
2 polypeptide of claim 1, wherein said nucleic acid molecule is operably linked to a
3 heterologous promoter.

1 4. An expression vector comprising a nucleic acid molecule, or its
2 complement, wherein the nucleic acid molecule encodes the polypeptide of claim 1.

1 5. A host cell comprising the expression vector of claim 4.

1 6. The host cell of claim 5, wherein the host cell is from a mammal.

1 7. A nucleic acid probe that specifically hybridizes with a nucleic acid
2 molecule encoding the polypeptide of claim 1.

1 8. The nucleic acid probe of claim 7, wherein the nucleic acid is a
2 DNA.

1 9. The nucleic acid probe of claim 7, wherein the nucleic acid is an
2 RNA.

1 10. An expression vector comprising a nucleic acid molecule, or its
2 complement, wherein the nucleic acid molecule selectively hybridizes to a sequence
3 selected from Table 1, wherein the hybridization reaction is incubated overnight at 37°C
4 in a solution comprising 40% formamide, 1 M NaCl and 1% SDS, and washed at 55°C in
5 a solution comprising 0.5x SSC.

1 11. An antibody that selectively binds to the polypeptide of claim 1.

1 12. The antibody of claim 11, wherein said antibody is a monoclonal
2 antibody.

1 13. The antibody of claim 11, wherein said antibody is a polyclonal
2 antibody.

1 14. An antisense polynucleotide comprising a sequence capable of
2 specifically hybridizing to a nucleic acid molecule encoding the polypeptide of claim 1.

1 15. A method for identifying a compound that modulates the
2 expression of a polypeptide in a cell, wherein said polypeptide has at least 80% amino
3 acid sequence identity to a polypeptide encoded by a nucleotide sequence selected from
4 the group consisting of the sequences set forth in Table 1, the method comprising the
5 steps of:

- 6 (a) culturing said cell in the presence of a modulator to form a first cell
7 culture;
- 8 (b) contacting RNA or cDNA from the first cell culture with a probe which
9 comprises a polynucleotide sequence encoding said polypeptide; and
- 10 (c) determining whether the amount of the probe which hybridizes to the
11 RNA or cDNA from the first cell culture is increased or decreased relative to the amount
12 of the probe which hybridizes to RNA or cDNA from a second cell culture grown in the
13 absence of said modulator.

1 16. A method for identifying a compound that modulates the
2 expression of at least two polypeptides in a cell, wherein each of said polypeptides has at
3 least 80% amino acid sequence identity to a polypeptide encoded by a nucleotide
4 sequence selected from the group consisting of the sequences set forth in Table 1, the
5 method comprising the steps of:

- 6 (a) culturing said cell in the presence of a modulator to form a first cell
7 culture;
- 8 (b) contacting RNA or cDNA from the first cell culture with at least two
9 probes, each probe comprising a polynucleotide sequence encoding one of said
10 polypeptides; and
- 11 (c) determining whether the amount of the probes which hybridizes to the
12 RNA or cDNA from the first cell culture is increased or decreased relative to the amount
13 of the probes which hybridizes to RNA or cDNA from a second cell culture grown in the
14 absence of said modulator.

1 17. A method for identifying a compound that modulates the activity of
2 a polypeptide, wherein said polypeptide has at least 80% amino acid sequence identity to
3 a polypeptide encoded by a nucleotide sequence selected from the group consisting of the
4 sequences set forth in Table 1, the method comprising the steps of:

5 (a) culturing cells expressing said polypeptide in the presence of a
6 modulator to form a first cell culture; and

7 (b) measuring the activity of said polypeptide or second messenger activity
8 in the first cell culture and determining whether the activity is increased or decreased
9 relative to the activity of said polypeptide or second messenger activity from a second cell
10 culture grown in the absence of said modulator.

1 18. A method for identifying a compound that modulates the activity of
2 at least two polypeptides, wherein each of said polypeptides has at least 80% amino acid
3 sequence identity to a polypeptide encoded by a nucleotide sequence selected from the
4 group consisting of the sequences set forth in Table 1, the method comprising the steps
5 of:

6 (a) culturing cells expressing said polypeptides in the presence of a
7 modulator to form a first cell culture; and

8 (b) measuring the activity of said polypeptides or second messenger
9 activity in the first cell culture and determining whether the activity is increased or
10 decreased relative to the activity of said polypeptides or second messenger activity from a
11 second cell culture grown in the absence of said modulator.

SEQUENCE LISTING

5

SEQ ID NO:1

189884

Cluster name: G protein-coupled receptor Ls189884 (putative GALR4 receptor)

SequenceID: LG610

10 Sequence: GGAGGGTACC TGCCCTCTGA TTCCCAGGAC TGGAGAACCA TCATCCCGGC
TCTCTGGTG GCTGTCTGCC TGGTGGGCTT CGTGGAAAC CTGTGTGTGA TTGGCATCCT
CCTTCACAAT GCTTGGAAAG GAAAGCCATC CATGATCCAC TCCCTGATTG TGAATCTCAG
CCTGGCTGAT CTCTCCCTCC TGCTGTITTC TGCACCTATC CGAGCTACGG CGTACTCCAA
AAGTGTITGG GATCTAGGCT GGTTTGTCTG CAAGTCCTCT GACTGGTTTA TCCACACATG
15 CATGGCAGCC AAGAGCCTGA CAATCGTTGT GGTGGCCAAA GTATGCTTCA TGTATGCAAG
TGGCCCAACC CAGCAAGTGG TTTTCAACT ACCCCATTG GTAATGGCGG TTGGCCTTTT
GACTGGGCT TACCTGTTA

SEQ ID NO:2

20 3098

Cluster name: Metabotropic glutamate receptor 6

SequenceID: NM_000843

Sequence: CGGAGGCCCG GGCAGGCCGG CTGAGCTAAC TCCCCAGAGC
CAAAGTGGAA GGCGCGCCCG GAGCGCCTTC TCCCCAGGAC
25 CCCGGTGTCC CTCCCCGCGC CCCGAGCCCG CGCTCTCCTT
CCCCCGCCCT CAGAGCGCTC CCCGCCCCCTC TGTCTCCCCG
CAGCCCGCTA GACGAGCCGA TGGCGCGGCC CCGGAGAGCC
CGGGAGCCGC TGCTCGTGGC GCTGCTGCCG CTGGCGTGGC
TGGCGCAGGC GGGCCTGGCG CGCGCGCGG GCTCTGTGCG
30 CCTGGCGGGC GGCCTGACGC TGGGCGGCCCT GTTCCCGGTG
CACGCGCGGG GCAGCGCGGG CCAGGCGTGC GGCCGCTGA
AGAAGGAGCA GGGCGTGCAC CGGCTGGAGG CCATGCTGTA
CGCGCTGGAC CGCGTCAACG CGAACCCGA GCTGCTGCC
GGCGTGCGCC TGGGCGCGCG GCTGCTGGAC ACCTGCTCGC
35 GGGACACCTA CGCGCTGGAG CAGGCGCTGA GCTTCGTGCA
GGCGCTGATC CGCGGCCCGCG GCGACGGCGA CGAGGTGGGC
GTGCGCTGCC CGGGAGGCCT CCCTCCGCTG CGCCCCGCGC
CCCCCGAGCG CGTCGTGGCC GTCTGGCGG CCTCGGCCAG
CTCCGTCTCC ATCATGGTCG CCAACGTGCT GCGCCTGTT
40 GCGATACCCC AGATCAGCTA TGCCCTCCACA GCCCCGGAGC
TCAGCGACTC CACACGCTAT GACTTCTTCT CCCGGGTGGT
GCCACCCGAC TCCTACCAGG CGCAGGCCAT GGTGGACATC
GTGAGGGCAC TGGGATGGAA CTATGTGTCC ACGCTGGCCT
CCGAGGGCAA CTATGGCGAA AGTGGGGTTG AGGCCTTCGT
45 TCAGATCTCC CGAGAGGCTG GGGGGGTCTG TATTGCCAG
TCTATCAAGA TTCCCAGGGA ACCAAAGCCA GGAGAGTTCA
GCAAGGTGAT CAGGAGACTC ATGGAGACGC CCAACGCCCG
GGGCATCATC ATCTTGCCA ATGAGGATGA CATCAGGCCGG
GTCCCTGGAGG CAGCTCGCCA GGCAACCTG ACCGGCCACT
50 TCCTGTGGGT CGGCTCAGAC AGCTGGGGAG CCAAGACCTC
ACCCATCTTG AGCCTGGAGG ACCTGGCCGT TGGGGCCATC
ACCATCCTGC CCAAAAGGGC CTCCATCGAC GGATTTGACC
AGTACTTCAT GACTCGATCC CTGGAGAACCA ACCGCAGGAA
CATCTGGTTC GCCGAGTTCT GGGAAAGAGAA TTTTAAGTGC
55 AAACTGACCA GCTCAGGTAC CCAGTCAGAT GATTCCACCC

GCAAATGCAC AGGCAGGAA CGCATCGGCC GGGACTCCAC
 CTACGAGCAG GAGGGCAAGG TGCAGTTGT GATTGATGCG
 GTGTATCCA TTGCCCACGC CCTCCACAGC ATGCACCAGG
 CGCTCTGCCA TGGGCACACA GGCTGTGCC CGGGATGGA
 5 ACCCACCGAT GGGCGGATGC TTCTGCAGTA CATTGAGCT
 GTCCGCTTCA ACGGCAGCGC AGGAACCCCT GTGATGTTCA
 ACGAGAACGG GGATGCGCC GGGCGGTACG ACATCTTCA
 GTACCAGGCG ACCAATGGCA GTGCCAGCAG TGGCGGGTAC
 CAGGCACTGG GCCAGTGGGC AGAGACCCCTC AGACTGGATG
 10 TGGAGGCCCT GCAGTGGTCT GGCGACCCCC ACGAGGTGCC
 CTCGTCTCTG TGCAGCCTGC CCTGCGGCC GGGGGAGCGG
 AAGAAGATGG TGAAGGGCGT CCCCTGCTGT TGGCACTGCC
 AGGCCTGTGA CGGGTACCGC TTCCAGGTGG ACGAGTTCAC
 ATGCGAGGCC TGTCTGGGG ACATGAGGCC CACGCCAAC
 15 CACACGGGCT GCCGCCAAC ACCTGTGGTG CGCCTGAGCT
 GGTCTCCCC CTGGGCAGCC CCGCCGCTCC TCCCTGGCGT
 GCTGGGCATC GTGGCCACTA CCACGGTGGT GGCCACCTC
 GTGCGGTACA ACAACACGCC CATCGTCCGG GCCTCGGGCC
 GAGAGCTCAG CTACGTCTC CTCACGGCA TCTTCCTCAT
 20 CTACGCCATC ACCTTCTCA TGGTGGCTGA GCCTGGGGCC
 GCGGTCTGTG CCGCCCGCAG GCTCTTCTG GGCTGGGCA
 CGACCTCAG CTACTCTGCC CTGCTCACCA AGACCAACCG
 TATCTACCGC ATCTTGAGC AGGGCAAGCG CTCGGTCACA
 CCCCCCTCCT TCATCAGCCC CACCTCACAG CTGGTCATCA
 25 CCTTCAGCCT CACCTCCCTG CAGGTGGTGG GGATGATAGC
 ATGGCTGGGG GCCCCGGCCC CACACAGCGT GATTGACTAT
 GAGGAACAGC GGACAGTGGA CCCCAGCAG GCGAGAGGGG
 TGCTCAAGTG CGACATGTG GATCTGTCTC TCATCGGCTG
 CCTGGGCTAC AGCTCCTGC TCATGGTCAC GTGCACAGTG
 30 TACGCCATCA AGGCCCCGTGG CGTGGCCGAG ACCTTCAACG
 AGGCCAACGCC CATCGGCTTC ACCATGTACA CCACCTGCAT
 CATCTGGCTG GCATTGTGC CCATCTTCTT TGGCACTGCC
 CAGTCAGCTG AAAAGATCTA CATCCAGACA ACCACGCTAA
 CCGTGTCTT GAGCCTGAGT GCCTCGGTGT CCCTCGGCAT
 35 GCTCTACGTA CCCAAAACCT ACGTCATCCT CTTCCATCCA
 GAGCAGAATG TGCAGAAGCG AAAGCGGAGC CTCAGGGCA
 CCTCCACGGT GGCAAGCCCC CCAAGGGCG AGGATGAGA
 GGCCCACAAAG TAGCAGGGCA GGTGGGAACG GGACTGCTT
 CTGCCTCTCC TTTCTTCTC TTGCTCGAG GTGGAAGCTG
 40 TATAGAGCCC GGGTCCACGG TGAACAGTCA GTGGCAGGG
 GTTTGCCAAG ACCATGCTCC GCGTCGGTGG GGCTGGCTT
 GAGAAGGAAC TGGACCCAGC TCTACCCGA TTCCAGCATG
 TGAGCTTCAT GCTTCTCAC CACAGACCA ACTCGCTTCC
 CATGGTGGGA AACAGGCCACC GAGAAGGTTC TAGCTCTAGA
 45 AAGGGACTAA ACTTATTCTC TCATCCGAAG TCCAAAGAGG
 ATGATGAAGC CCTGGGCTTT GCCTGGTTG CGGGAGATT
 CCTCCCTCA GTCAACCCCC ATAACCTGGG GATTGGGAG
 TGTGGAAGAA CGTAGACCC CCAGAATGAA ACATGGGGTT
 GGAGTGGAGG AGGAGCTGTC TCAGCAAGAG GAGACCTGGG
 50 GCTGTGCATC TGGATGGAGG CACTCAGGCC TGGTAGGAT
 TCCTCTGGCA CGGAGGGAGA GACCCCTGGT GAGACCCCTG
 TGAGCATGGG AAGGGCCTGC AGTGGCGCG GGAGTGAAGCT
 GAGGAACCTGG GGTGCGCCCATGAGATTC CCAATGCCAT
 GGGCTTCCC CCATCCCCCCC GGGATTGGGC AAGGTCAGAC
 55 TTAGAGTACA GCTGTTTCC TCCCCTCTGT GTACTCCCTT
 AAATCACCCCC AACCTTGGCC AGGCATGGTG GCTCACACCT
 GTAATCCCCAG CACTTGGGA GGCGAGGCA GGTGGATCAC
 CTGAGGTCCG GAGTCGAGA CCAGCCTGGC CAATGTGGTG
 AAACCCCTGTC TCTACTAAAA ATACAAAAAT TAGCCAGGTG
 60 TGATGGTGGG TGCCTGTAAT CCCAGTTACT TGGGAGGCTG

AGGCAGGAGA ATCGCTTGAA CCTGGGAGGT GGAGGTTGCA
 GTGAGCTGTG ATTGTGCCAC TGTACTCCAG CCTGGGTGAC
 AGAGCGAGAC TCTGTCTCAA AAAAACAAAA CAAAAAAACA
 CAAAAAAAAC CCCCAAACCT GAAGAAATTC AGATAACACGT
 5 GTGTAATGTT AGTGTATGTGA GAACAAGGAG CAGGGGTGCA
 TTTGTGTTGT GTCGGGGTTC GGGATGGGT TAGGAGCTCC
 AGGTTGGGAG CAGTGACAGA GAGTCATGGC CGTGGTGAGG
 GTGAATCCCA AGTGGATGGC TCAGGACGGG TATGGAAACC
 CTTCATTCCT CATAAGGTACT GGGAAAGTCCA TTTGCAAGCT
 10 GAGCGCCAGG CCTGGGGAGG AAGAGGCTTG GGCTGCAGAT
 GCACGCACAT TTGTTTTCA CTGATAGTTT TTACAAAAAG
 CTTGGTTAA GTTATGGAAT TTTATGTCCC TGGGAGTAGA
 ATTACATTT GTTAAATTGA CCACTGTITA AGATCAGTAT
 ACATTCTCTA GTCTGTGATG TCTGGAGCTA GTTTGAGGG
 15 TGAACCACAC TTTATCCAAC ATACAAACCT TCCCCTGCAG
 CTTCTCTGGT GCGCAGTTGG TTTGACCGT GGGACTAGGT
 GCTTCTGCAG GTTTTAAGTA ATTAACTTAA AAGCITCTCC
 TCTGAGAAAC ATTTCTGTG CGCTACTGAC TCTCCTCTC
 CACATTGTT GTGTTCTAG GGCTTCTCTA TAGTGCACAT
 20 TAGGACGTTT CATTGTTGC TGAATGCTT CCAGAATTAT
 TTATTCCATA GGGTTCTCT CCTGTGCAGC TCTCTCATGG
 GTAATGGGGC GTGTTTTCTT GCCAAAGGCG GTTCCACCC
 CGTGATTGTA TAGGGCTCTT CTCCTGTATG AACTCTGAGA
 TCAGTGAGCT CTGATCTCCA AGGGAAAGTT TTCCTGCATT
 25 TGCTGTTTC TCATGTCTCT CCCAGTGTGA ATTCTCTGGC
 TTCTAGCTGA AAACCTTTCC ACAGTTTAC ATTCTATGTGG
 TTTCTCCAC TGTGAACCTCT GTGATTCAAGA ATCAGAAGCA
 GTTCTTAGTA GAGGCATTTTC TACACTGATT GCACTGAGGA
 TATCTCCCCA GTGTGAAGTT TCTGGCATAG AGTCCCTGGCT
 30 TCCCAGAC GACTTCACA CTCTGCCATG TTCATGCCTG
 TGGGCCTCTC TGGCAGGAAC TCTGATGCAC CGCGAGGCC
 ATGTACTCT GTGGCTTCT CACATTGGT CTACTTGAG
 GGTATCTCCA CAGCATGCAC CATTCTGGT ACAGGGGGAC
 ATCCTCTGTT ACTGAAGATG TTGTCATATT TAGTACCTC
 35 ACAAGGTTTC TCTCCTCCA GAATTTCTG ATGTACACAA
 ATAACGTACT TCCACAAGAG GGCTTTCCA CACTCGGTGT
 GTGCATACAG TTTCTGCCCTG TGATCATTTC TTTATGTTAT
 TATTTTATT TTTGAGATA GGGTCTTGCT CAATTCTTA
 GGCTGGAGTG CAGTGGCACG ATCATAGCTC ACTGAAGTTT
 40 CGACCTGGGC TCAAGCAATC CTCCCGCTTC AGCCTCCTGA
 GTAGCTGGTG CGCACGACCA TACCCAGCTA ATGTTTTATT
 TTTTGTAGAG ACGAGGTCTC ACTATGTGTC CCAGGCTGGT
 CTCGAACCTTC TGAGCTCGAG CGATCCTCCT GCCTCCACCT
 CCCAAAGTGT TCGGATTACA AACGTGAGCC ATCGCACCTA
 45 GCCTCTTGA TCATTTCTGT GGTGTTCACT GGGGGTTGAC
 AGCTCCCTAA AGATTTCCCT GTTTTTTGC ATGCATGGGT

SEQ ID NO : 3

22315

50 Cluster name: G protein-coupled receptor GPR92

SequenceID: NM_020400

Sequence: ATGTTAGCCA ACAGCTCCTC AACCAACAGT TCTGTTCTCC
 CGTGTCCCTGA CTACCGACCT ACCCACCGCC TGCACTTGGT
 GGTCTACAGC TTGGTGCTGG CTGCCGGGCT CCCCCCTCAAC
 55 GCGCTAGCCC TCTGGGTCTT CCTGCGCGCG CTGCGCGTGC
 ACTCGGTGGT GAGCGTGTAC ATGTGTAACC TGGCGGCCAG
 CGACCTGCTC TTCACCCCT CGCTGCCCGT TCGTCTCTCC
 TACTACGCAC TGCAACCACTG GCCCTTCCCC GACCTCCTGT

GCCAGACGAC GGGCGCCATC TTCCAGATGA ACATGTACGG
 CAGCTGCATC TTCTGATGC TCATCAACGT GGACCGCTAC
 GCCGCCATCG TGCACCCGCT GCGACTGCGC CACCTGCGGC
 GGCCCCGCGT GGCGCGGCTG CTCTGCCCTGG GCGTGTGGGC
 5 GCTCATCCTG GTGTTGCCG TGCCC GCCGC CGCGTGCAC
 AGGCCCTCGC GTTCCGCTA CGGGACCTC GAGGTGCGCC
 TATGCTTCGA GAGCTTCAGC GACGAGCTGT GGAAAGGCAG
 GCTGCTGCCCTCCTCGTGC TGGCCGAGGC GCTGGGCTTC
 CTGCTGCCCTGGCGGCTG GGTCTACTCG TCGGGCCGAG
 10 TCTTCTGGAC GCTGGCGCGC CCCGACGCCA CGCAGAGCCA
 CGGGCGGCGG AAGACCGTGCGCCTCGT GGCTAACCTC
 GTCATCTTCC TGCTGTGCTT CGTGCCTAC AACAGCACGC
 TGGCGGTCTA CGGGCTGCTG CGGAGCAAGC TGGTGGCGGC
 CAGCGTGCCT GCCCCGCGATC GCGTGCCTGG GGTGCTGATG
 15 GTGATGGTGC TGCTGGCCGG CGCCAACCTGCGTGGACC
 CGCTGGTGTAC TACTTTAGC GCGGAGGGCT TCCGCAACAC
 CCTGCCGGC CTGGGCACTC CGCACCCGGC CAGGACCTCG
 GCCACCAACG GGACGCGGGC GGCGCTCGCG CAATCCGAAA
 GGTCCGCGGT CACCAACGAC GCCACCAGGC CGGATGCCGC
 20 CAGTCAGGGCTGCTCGAC CCTCCGACTC CCACCTCTCG
 TCTTCCTTCA CACAGTGTCC CCAGGATTCC GCCCTCTGA

SEQ ID NO: 4

30875

25 Cluster name: G protein-coupled receptor GPR87

SequenceID: NM_023915

Sequence: GGCACGAGGG TTTCTGTTTC ATGCTTTACC AGAAAATCCA
 CITCCCTGCC GACCTTAGTT TCAAAGCTTA TTCTTAATTAA
 GAGACAAGAA ACCTGTTCA ACTTGAAGAC ACCGTATGAG
 30 GTGAATGGAC AGCCAGCCAC CACAATGAAA GAAATCAAAC
 CAGGAATAAC CTATGCTGAA CCCACGCCCTC AATCGTCCCC
 AAGTGTTC TGACACGCAT CTTTGCTTAC AGTCATCAC
 AACTGAAGAA TGGGGTTCAA CTTGACGCTT GCAAATTAC
 CAAATAACGA GTCGACCGC CAAGAGAGTC ACAATTCAAGG
 35 CAACAGGAGC GACGGGCCAG GAAAGAACAC CACCCCTTCAC
 AATGAATTG ACACAATTGT CTTGCCGGTG CTTTATCTCA
 TTATATTGT GGCAAGCATC TTGCTGAATG GTTGTAGCAGT
 GTGGATCTTC TTCCACATTA GGAATAAAAC CAGCTTCATA
 TTCTATCTCA AAAACATAGT GGTTGCAGAC CTCATAATGA
 40 CGCTGACATT TCCATTTCGA ATAGTCCATG ATGCAGGATT
 TGGACCTTGG TACTTCAAGT TTATTCTCTG CAGATACACT
 TCAGTTTGT TTTATGCAAA CATGTATACT TCCATCGTGT
 TCCCTGGGCT GATAAGCATT GATCGCTATC TGAAGGGTGGT
 CAAGCCATTG GGGACTCTC GGATGTACAG CATAACCTTC
 45 ACGAAGGTTT TATCTGTTG TGTTGGGTG ATCATGGCTG
 TTTGTCTT GCCAAACATC ATCCTGACAA ATGGTCAGCC
 AACAGAGGAC AATATCCATG ACTGCTAAA ACTTAAAGT
 CTTTGGGGG TCAAATGGCA TACGGCAGTC ACCTATGTGA
 ACAGCTGCTT GTTTGTGGCC GTGCTGGTGA TTCTGATCGG
 50 ATGTTACATA GCCATATCCA GGTACATCCA CAAATCCAGC
 AGGCAATTCA TAAGTCAGTC AAGCCGAAAG CGAAAACATA
 ACCAGAGGAT CAGGGTTGTT GTGGCTGTGT TTTTACCTG
 CTTTCTACCA TATCACTTGT GCAGAAATTCC TTTTACTTT
 AGTCACCTAG ACAGGCTTT AGATGAATCT GCACAAAAAA
 55 TCCTATATTA CTGCAAAGAA ATTACACTT TCTTGTCTGC
 GTGTAATGTT TGCCTGGATC CAATAATTAA CTTTITCATG
 TGTAGGTCA TTTCAAGAAG GCTGTTCAAA AAATCAAATA
 TCAGAACCAAG GAGTGAAGC ATCAGATCAC TGCAAAGTGT

GAGAAGATCG GAAGTTCGCA TATATTATGA TTACACTGAT
GTGTAGGCCT TTTATTGTTT GTTGGAAATCG ATATGTACAA
AGTGTAAATA AATGTTCTT TTCATTATCC TTAAAAAAA AA

5 SEQ ID NO:5

54602

Cluster name: Pheromone receptor (PHRET) pseudogene

SequenceID AF253316

10 Sequence: TCTGACAGAC AACACCTTT TGCTTTCTT CCACATCTTC
ACACTCCTTC AGGATCAAAA ACCTAAGCCA CATGACTGGA
TGAGCCGTCA CTTGCCCTC ATTGGGTAG TGATGGCCT
CACTGTAGTG GATGTTTGC CTCCAGATAT GCTTGAATCA
CTGCATTTG GGAATAACTT CAAATGCAAG TCCTGATCT
AAATAAACAG AATGACGAAG GGCCTATGTT TCTATACCAC
15 CTGTCTCCTG AATATACACC AGGCCAGCAT AATCAGCCTC
AGCAACTTCT GGTTGGAAAG CTTAAACAT AAATTACAA
ATAACATTGT CAGTGCCTC TTTTTCTT TTTGTTCCCT
CAATTGTCT TTCAGTAGTG ACATAATAATT CTTCACTGTG
GCTTCTTCCA TTGTGACCCA GACCAATCTA CTTAAGGTCC
20 GCAAATACTG CTCACGTTCT CCCATGAAAT CCATCATGTG
GGGAGTGTGTT TCCTTGTAGG ATTACGCTGC TCTCAAGTGC
ATACATGATG ATCTTTTGT CCAAGCATCA GAAGTGTAC
CAGCATCTTC ACAGTACCAAG CCTTTCCCCA AGATCCTCGC
CAGAGAAAAG GGTTACCCAG ATCATCCTGC CACTGGTGAA
25 TTGCTTGTG GTCATGTTCT GGGTGGACCT TATCATCTCA
TCCTCTTCAT CCCTGTTATG GACGTATAAC CCAGTCATCC
TGAGCATCTA GAACCTGTT GCCTGTGCT ATGCCACTCT
CGTTCCATTG GTACAAATCC GCTCTGATAA AAGAATAGTC
AATATTCTCC AAAAATGGA ATTAAAGTGC TATAATTGTT
30 TAATGTGTTG GTGATGAAAA ATATTCTAA AAATTAGTCT
CATTCTATAG TAAATTGTT CAAGTAGCCC CAGATTAGC
TTACTGAGTT TAAATAAAAT GCGTGGAAATT ACACTTTAT
TATATTCTTA TGCTTCTGAA ACTGAGGCAT CTAAGGACTA
TGTAGTTCT TCAGTCAAT GTTCAACCATA GATTGACATT
35 TCAGATATCA AGTCTTTGC ACTTTTATT TTATGTTAAC
TTTGTACAAG AAAATGTTTC TCTCTTTTG AAGTACATT
TTAAAAAAATT TGTTTGGTA TCAATCTCTC AATGTTTTA
CTTTGAAAA TATTACTTA CTCTGTTAT GAATGATACT
TTAGCTCAAT ATTCAATTCT AGCTTTAAG CCATGCTTGC
40 TCATTGTACC TCCCTGACTA AAAAATTA TGTCTATTG
GATTAAAT TTAATCTAGA ATTCAATTAA ACG

SEQ ID NO:6

55728

45 Cluster name: ETL protein

SequenceID: NM_022159

50 Sequence: GTGAAATTAA AACTCCAGTC CTGTGGCGAA AATGCTAATT
GCACTAACAC AGAAGGAAGT TATTATTGTA TGTGTGTACC
TGGCTTCAGA TCCAGCAGTA ACCAAGACAG GTTATCACT
AATGATGGAA CCGTCTGTAT AGAAAATGTG AATGCAAAC
GCCATTAGA TAATGTCTGT ATAGCTGAA ATATTAATAA
AACTTTAACAA AAAATCAGAT CCATAAAAGA ACCTGTGGCT
TTGCTACAAG AAGTCTATAG AAATTCTGTG ACAGATCTT
55 CACCAACAGA TATAATTACA TATATAGAAA TATTAGCTGA
ATCATCTTCAT TTACTAGGTT ACAAGAACAA CACTATCTCA

GCCAAGGACA CCCTTTCTAA CTCAACTCIT ACTGAATTG
TAAAAACCGT GAATAATTIT GTTCAAAGGG ATACATTGT
AGTTTGGGAC AAGTTATCTG TGAATCATAG GAGAACACAT
5 CTTACAAAAC TCATGCACAC TGTTGAACAA GCTACTTAA
GGATATCCCCA GAGCTTCCAA AAGACCACAG AGTTTGATAC
AAATTCAACG GATATAGCTC TCAAAGTTT CTTTTTGAT
TCATATAACA TGAAACATAT TCATCCTCAT ATGAATATGG
ATGGAGACTA CATAAAATATA TTTCAGAAAGA GAAAAGCTGC
ATATGATTCA AATGGCAATG TTGCAGTTGC ATTGTTATAT
10 TATAAGAGTA TTGGTCCTT GCTTTCATCA TCTGACAAC
TCTTATTGAA ACCTCAAAAT TATGATAATT CTGAAGAGGA
GGAAAGAGTC ATATCTTCAG TAATTTCACT CTCATGAGC
TCAAACCCAC CCACATTATA TGAACCTGAA AAAATAACAT
15 TTACATTAAG TCATCGAAAG GTCACAGATA GGTATAGGAG
TCTATGTGCA TTTTGGATT ACTCACCTGA TACCATGAAT
GGCAGCTGGT CTTCAGAGGG CTGTGAGCTG ACATACTCAA
ATGAGACCCA CACCTCATGC CGCTGTAATC ACCTGACACA
TTTGCAATT TTGATGTCT CTGGTCCTTC CATTGGTATT
20 AAAGATTATA ATATTCTTAC AAGGATCACT CAACTAGGAA
TAATTATTTC ACTGATTGT CTTGCCATAT GCATTITAC
CTTCTGGTTC TTCAGTGAAGA TTCAAAGCAC CAGGACAACA
ATTACACAAAA ATCTTGCTG TAGCCTATT TTTGCTGAAC
TTGTTTTCT TTGTTGGGATC AATACAAATA CTAATAAGCT
25 CTTCTGTCA ATCATTGCCG GACTGCTACA CTACTTCTT
TTAGCTGCTT TTGCAATGGAT GTGCATTGAA GGCATACATC
TCTATCTCAT TTGTTGGGT GTCATCTACA ACAAGGGATT
TTTGACAAAG AATTTTATA TCTTTGGCTA TCTAAGCCCA
GCCGTGGTAG TTGGATTTC GGCAGCACTA GGATACAGAT
ATTATGGCAC AACCAAAGTA TTGTTGGCTTA GCACCGAAAAA
30 CAACCTTATT TTGAGTTTA TAGGACCAGC ATGCCTAATC
ATTCTTGITA ATCTCTTGGC TTTTGGAGTC ATCATATACA
AAGTTTTTCG TCACACTGCA GGGTTGAAAC CAGAAGTTAG
TTGCTTTGAG AACATAAGGT TTGTTGCAAG AGGAGCCCTC
GCTCTTCTGT TCCTTCTCGG CACCACCTGG ATCTTTGGGG
35 TTCTCCATGT TGTGACGCA TCAGTGGTTA CAGCTTACCT
CTTCACAGTC AGCAATGCTT TCCAGGGGAT GTTCATTTC
TTATTCTGT GTGTTTATC TAGAAAGATT CAAGAAGAAT
ATTACAGATT GTTCAAAAAT GTCCCCTGTT GTTTGGATG
40 TTTAAGGTTAA ACATAGAGAA TGGTGGATAA TTACAACACTGC
ACAAAAATAA AAATTCCAAG CTGTGGATGA CCAATGTATA
AAAATGACTC ATCAAATTAT CCAATTATTA ACTACTAGAC
AAAAAGTATT TAAATCACT TTTCTGTGTT ATGCTATAGG
AACTGTAGAT AATAAGGTA AATTATGTAT CATATAGATA
TACTATGTT TTCTATGTGA AATAGTTCTG TCAAAAATAG
45 TATTGCAGAT ATTGGAAAG TAATTGGTTT CTCAGGAGTG
ATATCACTGC ACCCAAGGAA AGATTTCTT TCTAACACGA
GAAGTATATG AATGTCCTGA AGGAAACAC TGGCTTGATA
TTTCTGTGAC TCGTGTGAC TGTGAAACTA GTCCCCTACC
ACCTCGGTA TGAGCTCCAT TACAGAAAGT GGAACATAAG
50 AGAATGAAGG GGCAGAAATAT CAAACAGTGA AAAGGGAATG
ATAAGATGTA TTTGAATGA ACTGTTTTT CTGTAGACTA
GCTGAGAAAT TGTTGACATA AAATAAAGAA TTGAAGAAC

SEQ ID NO: 7

55 160221

Cluster name: G Protein-Coupled Receptor GPR27

SequenceID: NM_018971

Sequence: ATGGCGAACG CGAGCGAGCC GGGTGGCAGC GGCGGCGGCG

AGGCAGGCCGC CCTGGGCCTC AAGCTGGCCA CGCTCAGCCT
 5 GCTGCTGTGC GTGAGCCTAG CGGGCAACGT GCTGTTCGCG
 CTGCTGATCG TGCAGGGAGCG CAGCCTGCAC CGCGCCCCGT
 ACTACCTGCT GCTCGACCTG TGCCCTGGCCG ACAGGGCTGCG
 CGCGCTCGCC TGCCTCCCGG CCGTCATGCT GGCGGCGCGG
 CGTGCAGCGG CGCGCGCGGG GGCGCCCGG GGCGCGCTGG
 GCTGCAAGCT GCTCGCCTTC CTGGCCGCGC TCTTCTGCTT
 CCACGCCGCC TTCCCTGCTGC TGGGCGTGGG CGTCACCCGC
 TACCTGGCCA TCGCGCACCA CCGCTTCTAT GCAGAGCGCC
 10 TGGCCGGCTG GCCGTGCGCC GCCATGCTGG TGTGCGCCGC
 CTGGGCGCTG GCGCTGGCCG CGGCCTTCCC GCCAGTGCTG
 GACGGCGGTG GCGACGACGA GGACGCGCCG TGCGCCCTGG
 AGCAGCGGCC CGACGGCGCC CCCGGCGCGC TGGGCTTCC
 GCTGCTGCTG GCCGTGGTGG TGGGCGCCAC GCACCTCGTC
 15 TACCTCCGCC TGCTCTTCTT CATCCACGAC CGCCGCAAGA
 TGCAGGCCGC GCGCCTGGTG CCCGCGTCA GCCACGACTG
 GACCTTCCAC GGCCCGGGCG CCACCGGCCA GGCGGCCGCC
 AACTGGACGG CGGGCTTCGG CGCGGGGCC ACGCCGCC
 CGCTTGTGGG CATCCGGCCC GCAGGGCCGG GCCGCGGCGC
 20 GCGCCGCCCTC CTCGTGCTGG AAGAATTCAA GACGGAGAAAG
 AGGCTGTGCA AGATGTTCTA CGCCGTACAG CTGCTCTTCC
 TGCTCCTCTG GGGGCCCTAC GTCGTGGCCA GCTACCTGCG
 GGTCTGGTG CGGCCCGCG CGTCCCCCA GGCCTACCTG
 ACGGCCTCCG TGTGGCTGAC CTTCGCGCAG GCCGGCATCA
 25 ACCCGCTCGT GTGCTTCCTC TTCAACAGGG AGCTGAGGGGA
 CTGCTTCAGG GCCCAGTTCC CCTGCTGCCA GAGCCCCCGG
 ACCACCCAGG CGACCCATCC CTGCGACCTG AAAGGCATTG
 GTTATGA

30 SEQ ID NO: 8

160314

Cluster name: G protein-coupled receptor Ls160314

SequenceID: ENSMDNA221753

Sequence: ATGAAGATCA AATATGACTT CCTATATGAA AAGGAACACA
 35 TCTGCTGCTT AGAAGAGTGG ACCAGCCCTG TGCAACAGAA
 GATCTACACC ACCTTCATCC TTGTCATCCT CTTCCCTCTG
 CCTCTTATGG TGATGCTTAT TCTGTACAGT AAAATTGGTT
 ATGAACCTTG GATAAAGAAA AGAGTTGGGG ATGGTTCACT
 GCTTCGAACCTT ATTCAATGGAA AAGAAATGTC CAAAATAGCC
 40 AGGAAGAAGA AACGAGCTGT CATTATGATG GTGACAGTGG
 TGGCTCTCTT TGCTGTGTGC TGGGCACCAT TCCATGTTGT
 CCATATGATG ATTGAATACA GTAATTGAA AAAGGAATAT
 GATGATGTCA CAATCAAGAT GATTTTGCT ATCGTGCAAA
 TTATTGGATT TTCCAACCTCC ATCTGTAATC CCATTGTCTA
 45 TGCATTTATG AATGAAAATCTAAAAAAATGTTTGCT
 GCAGTTGTT ATTGCAATGTT AAATAAAACC TTCTCTCCAG
 CACAAAGGCA TGGAAATTCA GGAATTACAA TGATGCGGAA
 GAAAGCAAAG TTTCCCTCA GAGAGAAATCC AGTGGAGGAA
 ACCAAAGGAG AAGCATTCAAG TGATGGCAAC ATTGAAGTCA
 50 AATTGTGTGA ACAGACAGAG GAGAAGAAAA AGCTCAAACG
 ACATCTTGCT CTCTTAGGT CTGAACGTGGC TGAGAATTCT
 CCTTTAGACA GTGGCATTAA A

SEQ ID NO: 9

55 160324

Cluster name: G protein-coupled receptor GPR86

SequenceID: NM_023914

Sequence: AACAGTATTTCCTTTCAA CACATCTATT GAAAGTGTG
GATAAATGCA GGATGTTAAT ATGCTATAAA CATAAAGTCT
GTTTTAAAAA AATAGCATTG AAAATCATG AAGGGCTTT
5 TGTTTCTTT TGTTGTATA TATGTTTATT GGTAACAGGT
GACACTGGAA GCAATGAACA CCACAGTGAT GCAAGGCTTC
AACAGATCTG AGCGGTGCC CAGAGACACT CGGATAGTAC
AGCTGGTATT CCCAGCCCTC TACACAGTGG TTTCCTTGAC
CGGCATCCTG CTGAATACTT TGGCTCTGTG GGTGTTGTT
10 CACATCCCCA GCTCCTCAC CTTCATCATC TACCTCAAAA
ACACTTTGGT GGCGACTTG ATAATGACAC TCATGCTTCC
TTTCAAAATC CTCTCTGACT CACACCTGGC ACCCTGGCAG
CTCAGAGCTT TTGTGTGTCG TTTCCTTCG GTGATATTTT
ATGAGACCACAT GTATGTGGGC ATCGTGTGT TAGGGCTCAT
15 AGCCTTGAC AGATTCCTCA AGATCATCAG ACCTTGAGA
AATATTTTC TAAAAAAACC TGTTTTGCA AAAACGGTCT
CAATCTTCAT CTGGTTCTTT TTGTTCTTCA TCTCCCTGCC
AAATATGATC TTGAGCAACA AGGAAGCAAC ACCATCGTCT
GTGAAAAAGT GTGCTTCCTT AAAGGGGCCT CTGGGGCTGA
20 AATGGCATCA AATGGTAAAT AACATATGCC AGTTTATTTT
CTGGACTGTT TTATCCTAA TGCTTGTT TTATGTGGTT
ATTGCAAAAA AAGTATATGA TTCTTATAGA AAGTCCAAAA
GTAAGGACAG AAAAAACAAAC AAAAGCTGG AAGGCAAAGT
ATTGTTGTC GTGGCTGTCT TCTTGTGTG TTTCCTCCA
25 TTTCATTTCG CCAGAGTTCC ATATACTCAC AGTCAAACCA
ACAATAAGAC TGACTGTAGA CTGCAAAATC AACTGTTTAT
TGCTAAAGAA ACAACTCTCT TTTCGGCAGC AACTAACATT
TGTATGGATC CCTTAATATA CATATTCTTA TGAAAAAAAT
TCACAGAAAA GCTACCATGT ATGCAAGGGGA GAAAGACCAC
30 AGCATCAAGC CAAGAAAATC ATAGCAGTCA GACAGACAAAC
ATAACCTTAG GCTGACAAC GTACATAGGG TTAACCTCTA
TTTATTGATG AGACTTCCGT AGATAATGTG GAAATCAAAT
TTAACCAAGA AAAAAAGATT GGAACAAATG CTCTCTTACA
TTTATTATC CTGGTGTACA GAAAAGATTA TATAAAATTT
35 AAATCCACAT AGATCTATTCA ATAAGCTGAA TGAACCATTAA
CTAAGAGAAAT GCAACAGGAT ACAAAATGCC ACTAGAGGTC
ATTATTCTTCTT TTTTTTTTTT AATTCAAGA
GCATTTCACT TTAACATTTT GGAAAGACT AAGGAGAAAC
GTATATCCCT ACAAAACCTCC CCTCCAAACA CCTCTCACA
40 TTCTTTCCA CAATTACACAT AACACTACTG CTTTGTGCC
CCTTAAATGT AGATATGTGC TGAAAGAAAA AAAAAACGCC
CAACTCTTGA AGTCCATTGC TGAAAACCTGC AGCCAGGGGT
TGAAAGGGAT GCAGACTTGA AGAGTCTGAG GAACGTGAAAGT
GGGTCAAGCAA GACCTCTGAA ATCCGGTAA AAGGATTTTC
45 TCCCTACAAT TACAAACAGC CTCTTCACA TTACAATAAT
ATACCATAGG AGGCACAAGC ACCATTATTA AGCCACTTGT
CTTACACCTT AAGTGTGTAC ATTCAAGTG TGAGAATGCT
GTGTTAACTA TTCTTGGAA TTCTCTTCT GTCCAGCAA
TACTCTAATG ATGGTAAAC ATGGCACCTA CTCAGCAATG
50 CCTTCCTTGA CCACAAACCC TATCCCCCTG CCCCACCCCTC
CTCATTAAAA ACAAAACTT CTACTGTTG GGTGTGTGAT
AGGGTTCTCA ATGCAGATCT CCTTTCTA GTAGCTATA
TTCTTGACTG CATCCGCTAA AAATGTTAA GCTCTTGAG
AGACAGACAT GCCAGATTT CTGGTATCT CCCATAATAC
55 GACCTACAGT CCATGGTCTA CAGATGTTT AAATAGAATT
GCTATTCTCG ATACATACAA AGACGTAATT GCTGACCCAC
AATCAGTAAC ATCCATATTG GGAGATTTT CAAAGGATGG
TGACCCCTGCT TGTATTATT TACCTGGTA TTTCCTTG
CATCCCTCTG TGATTCAAAA AAGTAAAATG TGGCTTTCTG
60 AAATGATGGA TAAGAGTCTA CATCTTCTAG AAAAAATACA

TAAAGGAGTA GTTAAGCTCT GTAAATGTGC CACGAGCTCC
AACACGACCA TCGTAGGGTG AAGCCCACGT TTTCTTCCAT
GCCCTCAAAG GCCCTAGAAC TTGCCTACCT TTCTGGCCTT
ACCTCCTAGC TACTTATCCA TCTCTGAAC TTTATACTCT
5 TGTATAAATT TCTAACTTTC AGAAAATGCC ATACTCTGTT
TTGGCACCAAC ACATGTATAT TTCCCCCTGG TACACTTGGA
AGACTCTTAT CCATCTGTGA AACCCCTATGT TGTCATCACT
TGGTCCATGA AATATTACCT GGCCAATATC CCACCACAC
CTCAAACCCA ATCACCCCCCT CCTCTGTATG CTGTACACACC
10 TATATTATTA AACTTATCAC ATTGCATTGT AATTACTTCC

SEQ ID NO:10

160458

Cluster name: G protein-coupled receptor Ls160458

15 SequenceID: AI733823

Sequence: TTAAATTAA AAAACTTAT TGGAATAGCA TGTTAGCAGC
AGTGAACAGG GCATGGACA GAAGGTTCC AAAACAAGTT
TAGCATGAAG GATGCCATAT GCTGTTGCCA ACAACTAGAA
CACGGTGAAT AAAGACACAG TTCTGAATGT CCAGCACAAC
20 CTCTGGCCTG CAACTATGTT CAGTGTGAT GATAAAACAAG
GTGGTGACTT GGAAGGAATC CCTATGTCAA GTGAGAAAAAA
AAAATGATGT CTGACCTCCT TATATATGTA AAAAATATAC
CTTCAGAGTC CGTCAGTAAG CTGGAAGAAG TGGATGTTGA
AGTTTTAAAC ATCGATGATG GGTCTCCAGT TGTTCATCAA
25 CCCATGGTGA AATAGCTGAA CGGTTCTGAA TCAAAGGTGA
TCCTAATAGT GAAGACATTA ACATTGAGA AAAAGTGCCT
ACAGATTATA TGGTGAATAAC ACGTGATGGG CTTCTTGAAG
GACTAGAGCA GTGTGTATTG AAAACAGAAC AAGAAATCAC
GTCAGTTAT

30

SEQ ID NO:11

160833

Cluster name: 5-HT5B receptor

SequenceID: AJ308679

35 Sequence: CCCCCCTCCAC GCCCGCACCT GCCCGGTCCA CGCCGAACTC
ACTGAGGACT CGTGTGCCCC CTGCCCTGGA GCTGCGATCC
CAAGCGCCGT GGAGGCCGCT AGCCTTCAG TGGCCACCGC
CGGCGTTGCC CTTGCCCTGG GACCCGAGAC CAGCAGCAGG
ACCCGGGACCC CCAAGCCGA GAGGGATACT CGGTTCGACC
40 CCGAGCGGCG CCGTCCTGCC GGGCGAGGG CGGCCCTTCT
CTGTCTTCAC GGTCTCTGGTG GTGACGCTGC TAGTGCTGCT
GATCGCCGCC ACTTCCCTGT GGAACCTGCT GGTTCCGGTC
ACCATCCCACGGC GGGTCCGTGC CTTCCACCCG GTGCCGCATA
ACTTGGTGGC CTCGACGGCC GTCTCGGACG AACTAGTGGC
45 AGCGCTGGCG ATGCCACCGA GCCTGGCGAG TGAGCTGTGCG
ACCGGGCGAC GTCGGCTGCT GGGCCGGAGC CTGTGCCACCG
TGTGGATCTC CTTCGACGCC GGAGCCTGTG CCACGTGTGG
ATCTCCCTCC ACGGCTGTGC TGCCCCGCCG GCCTCGGGAA
CGTGGCGGCC ATGCCCTGG GCGCGACGG GGCCATCACA
50 CGGCACCTGC AGCACACGCT GCGCACCTGC AGCCGCGCCT
CGTTGCTCAT GATCGCGCTC ACCCGGGTGC CGTCGGCGCT
CATGCCCTC GCGCCGCTGC TCTTGGCCG GGGCGAGGTG
TGCAGCGCTC GGCTCCAGCG CTGCCAGGTG AGCCGGGAAC
CCTCCTATGC CGCCTCTCC ACCCGCGCGC CCTTCCACCT
55 CGCGCTTGGC GTGGTGCCTGTTGCTACCG GAAGATCTAC

GAGGCGGCCA AGITTCGTTT CGGCCGACGC CGGAGAGCTG
TGCTGCCGTT GCGGCCACC ATGCAGGTGA GGGGTGGGCT
GAGGAACGTT GCTTGGCGA AGCGGTTGCT AGAGAAAGGAG
GCGGCTTCGC GAATGGC

5

SEQ ID NO:12

162615

Cluster name: G protein-coupled receptor Ls162615

SequenceID: BF115152

10 Sequence: TTGAAGCCAC TGAGACATT TIGTTTATT CCCAGACCCC
TAAATCAGAA AACCCGATCG AATACTGAGC ATAATTCTT
CATTGACATT TGTCTCTAAA TGTCAAGITG TTCTGGAAAT
TTTTCTTGA TTTTNGATT CATTGCCCTA TTCATTGAG
ACAAACTGAG TTAGCATGAT GTTGTGGAG GAATCTCCAG

15 TATGAGAAAAA TGCATAATGG CCTTTGTTT GCAGTGGGTT
GAAAGGCTTT GAGAATTGG GTTGGCAGA TAAATCTGAT
GAGTTTGCT TTTCTGTTG CTTCCAAGAA CTTAAGGCAG
ACAACCTGTT GAACAGAAGT TGTGCGAGCT TACTGTCCAA
GAGTATTCCA AAGCATAAGA TAAAAAAATCC CTGGAATGCA

20 TTGAGTAAAG CAAAAATAAC ATGCCAAGCC AGATTCTGGC
TGTCCACTAT TGTTCTTATT CCAAAGCCCC AGGTGAGCCC
TAGCAGAGGG GTCAGAATGA GGAGGCTT CCCCACGCAG
ATGATGGTGG CCTTGTACATC CCCACTCAGT CTTCCCCAA
CAGTCGGCCT

25

SEQ ID NO:14

189874

Cluster name: Neuromedin U receptor 2

SequenceID: NM_020167

30 Sequence: ATGGAAAAAC TTCAGAACATGC TTCTGGATC TACCAGCAGA
AACTAGAAGA TCCATTCCAG AAACACCTGA ACAGCACCGA
GGAGTATCTG GCCTTCCTCT GCGGACCTCG GCGCAGCCAC
TTCTTCCTCC CCGTGTCTGT GGTGTATGTG CCAATTGGT
TGGTGGGGGT CATTGGCAAT GTCTGGTGT GCCTGGTGTAT

35 TCTGCAGCAC CAGGCTATGA AGACGCCAC CAACTACTAC
CTCCTCAGCC TGGCGGTCTC TGACCTCTG GTCTGCTCC
TTGGAATGCC CCTGGAGGTC TATGAGATGT GGCGCAACTA
CCCTTTCTTG TTCGGGCCCCG TGGGCTGCTA CTTCAAGACG
GCCCTCTTG AGACCGTGTG CTTCGCTCC ATCCTCAGCA

40 TCACCACCGT CAGCGTGGAG CGCTACGTGG CCATCCTACA
CCCGTTCCCG GCCAAACTGC AGAGCACCCG GCGCCGGGCC
CTCAGGATCC TCGGCATCGT CTGGGGCTTC TCCGTGCTCT
TCTCCCTGCC CAACACCAGC ATCCATGGCA TCAAGTTCCA
CTACTTCCCC AATGGGTCCC TGGTCCCAGG TTCGGCCACC

45 TGTACGGTCA TCAAGCCCAT GTGGATCTAC AATTCTCATCA
TCCAGGTAC CTCCTTCCTA TTCTACCTCC TCCCCATGAC
TGTCTCATCGT GTCTCTACT ACCTCATGGC ACTCAGACTA
AAGAAAGACA AATCTCTTGA GGCAGATGAA GGGATGCAA
ATATTCAAAG ACCCTGCAGA AAATCAGTCA ACAAGATGCT

50 GTTGTCTTG GTCTTAGTGT TTGCTATCTG TTGGGCCCCG
TTCCACATTG ACCGACTCTT CTTCAGCTT GTGGAGGAGT
GGAGTGAATC CCTGGCTGCT GTGTTCAACC TCGTCCATGT
GGTGTCAAGGT GTCTCTTCT ACCTGAGCTC AGCTGTCAAC
CCCATTATCT ATAACCTACT GTCTCGCCGC TTCCAGGCAG

55 CATTCCAGAA TGTGATCTCT TCTTCCACA AACAGTGGCA

5 CTCCCAGCAT GACCCACAGT TGCCACCTGC CCAGCGGAAC
ATCTTCCTGA CAGAAATGCCA CTTTGTGGAG CTGACCGAAG
ATATAGGTCC CCAATTCCCA TGTCAAGTCAT CCATGCACAA
CTCTCACCTC CCAACAGCCC TCTCTAGTGA ACAGATGTCA
AGAACAAACT ATCAAAGCTT CCACTITAAC AAAACCTGA

SEQ ID NO:15

189876

Cluster name: G protein-coupled receptor Ls189876

10 SequenceID: ENSMDNA207850

Sequence: ATGAACCAGA CTTTGAATAG CAGTGGGACC GTGGAGTCAG
CCCTAAACTA TTCCAGAGGG AGCACAGTGC ACACGGCCTA
CCTGGTGCTG AGCTCCCTGG CCATGTTAC CTGCCTGTGC
GGGATGGCAG GCAACAGCAT GGTGATCTGG CTGCTGGGCT
15 TTCGAATGCA CAGGAACCC TTCTGCATCT ATATCCTCAA
CCTGGCGGCA GCCGACCTCC TCTTCTCTT CAGCATGGCT
TCCACGCTCA GCCTGGAAAC CCAGCCCCG GTCAATACCA
CTGACAAGGT CCACCGAGCTG ATGAAGAGAC TGATGTACTT
TGCCTACACA GTGGGCCTGA GCCTGCTGAC GCCATCAGC
20 ACCCAGCGCT GTCTCTCTGT CCTCTTCCCT ATCTGGTTCA
AGTGTCAACCG GCCCAGGCAC CTGTCAAGCCT GGGTGTGTGG
CCTGCTGTGG ACACTCTGTC TCCTGATGAA CGGGTTGACC
TCTTCCTCT GCAGCAAGTT CTTGAAATTCA AATGAAGATC
GGTGCTTCAG GGTGGACATG GTCCAGGCCG CCCTCATCAT
25 GGGGGTCTTA ACCCCAGTGA TGACTCTGTC CAGCCTGACC
CTCTTGTCT GGGTGCAGGAG GAGCTCCAG CAGTGGCGGC
GGCAGCCCAC ACGGCTGTT GTGGTGGTCC TGGCCTCTGT
CCTGGTGTTC CTCATCTGTT CCCTGCCTCT GAGCATCTAC
TGGTTGTGC TCTACTGGTT GAGCCTGCCG CCCGAGATGC
30 35 AGGTCCCTGTG CTTCAGCTTG TCACGCCCT CTCCTGCCGT
AAGCAGCAGC GCCAACCCCG TCATCTACTT CCTGGTGGGC
AGCCGGAGGA GCCACAGGCT GCCCACCAAGG TCCCTGGGG
CTGTGCTCCA ACAGGGCCTT CGCGAGGAGC CCGAGCTGGA
AGGTGGGAG AGGCCAACCG TGGCACCAA TGAGATGGGG GCTTGA

SEQ ID NO:16

189881

Cluster name: G protein-coupled receptor Ls189881

SequenceID: ENSMDNA136950

40 Sequence: ATGACCCAAC TTGGAAATGA CATTCCAAG ACCACAAATG
ACATTCCAAGTACCAAGGAT GTCTCTATGC CCAGTGTGG
GGCCACACCA GATGCCGAGG CCTCTCCACC CCAGGAGGGC
TGCCTCCCTCC TCCTAGGTGA CAATGAAGAA TGTACTGTC
AGTCACTGGG CTCAGTGGTC GTCTCTGGC ATGAGCTGGG
45 TTTCATGAG CTCAGGAATG GGAAGCATGA CTCTGCCCT
GAGGCCACAT GCCACCTCCA TAGCGGATCT TTCTTCTGG
CTGGAGGGGA AGTCACTTCT TCCCATGAAA CTATTTATC
TATAAAATCTC CTCTCCTTGT TGGAGACCAA AGCCCAGCTG
CTCTGCTTG GTGCCCTGGT GGCTGGGGG CTCAAGGAGT
50 55 CTCAGAACCT CAAGGTCTGG AGCAGCCCCCT ATGTGACCTA
CATCCTTAAC CTGGCCACTG TTGATATGGT CAACCTCTCC
TGTGTAACTG TGATCCTGCT GGAGAAAATC CTCAATGCTGT
ATCACCAGGC GGCATTGCAG GTGGCTGTGT TTCTGGATCC
TGTCTCCTAT TTCTCCGACA CAGTGGGTCT CTGTCTCCTG
GTGGCCATGA GTATTGAGAG CTTTCTCTGT GCCCTCTGTC

CCACCTGGTG CTGCCACCGC CCAGAGCACA CCTCTGCCAT
 5 GGCCTATCT CAAAATATTG TCACATTCA GGTAGGACT
 TTAGCCCGTG AAGTTGGAT GCCTGGAAGT AAGAGGCAGG
 TTGATCTCAC AGAGTTGGC TGCTGCTATG TTCAGGCAGG
 GGATACAATT TGGCATTT ATGTGCCCTT ACCCTGGGCC
 AACAGTTCCC TTGGAGTGAT TTCACTGCTG CTGGTTTCA
 CCATGATTGT GGACCGITGG TTTTAAGAG CTGAGGAGGA
 AGGAACAGGA GTGGAACAG ITAAACATC ACAGAGCTCA
 10 CTGTTCTTAT CAAGATTCA GCTATTATTCT TGA

10

189884

SEQ ID NO:17

189883

Cluster name: G protein-coupled receptor Ls 189883

15 SequenceID: ENSMDNA163742

Sequence: ATGTTGCTGT GCTCTCTGCT TCCCGCCCTT GTGGGATCTC
 TCTCTGGGGC TGCTGTTCT GGCCTAATAG GCTGGCGGTT
 20 GCCAGGGAAG AGCCCCCGCT TTGACTGTCC AGGGGATGTG
 GTGGTCAGGG CCAGCTTCTC CATCTTCCAC CTGTACAACA
 TCACCCCTGTT TGATTTCACT GCTCCACCAG CTGGCTTGGA
 25 GTCTCAAGC GTTCCACCT GGGGCTACTG GGAAGCCCAA
 GGATTACAT TTGCCATGGA GGAGATCAAC AGGGACGCC
 ACCTGCTCCC CAGCCTCAGG CTGGGCTTCT CCATCCGGAA
 CTCTGGGCTG GGTATACTG CGCTGTGGGA GGCCAAGGTC
 30 AGCCCTCCT CCACACTGGC CAGCCTCAGC GACAGGACCC
 AGTTCCCATC CTTCTTCACT ACCCTGCTCA GTCACCTCAC
 GACCACCCAT GCAGTGGTGC AGCTGATGCT TCACCTCCGA
 TGGTCTGGG TGAGCGTCTT GGCGCAGGGG GACGACTTTG
 AGCTGCAGGG CAGGTCTCTG GTCGTCCAGG AGCTGGGCCA
 35 GGCTGGGTC TGCAATTGAAT TCCAACCTTG CATCCCCACC
 CGGGAGTCCC TGAAGATGAA AAACATCATC TGGCTGATGG
 AGAACTGTAC GGCCACCATC ATCCTGAAGG AAAGCAAAGT
 ACACATCGCC TACACAGTGG TCTATGCCAT CGCCCAGGCC
 CTGGCAGGCT GCAAGCATGG GGACCAGGGG TGTGCCGATG
 CCTGGGACTT CCAGCCCTGG CTGCTGCTTC GTCCCTCTCAA
 GAACGTGCAT TTCAAGACCC CTGATGGGAC AGAGATCATG
 TTTGATGCCA ACGGAGATT ATTACAGAA TTGATGTTG
 TCTATGGACA GAAGACCACT GAGGGCTGA

40 SEQ ID NO:18

LS_ID 189884

Cluster name: G protein-coupled receptor Ls189884

SequenceID: ENSMPRT108574

Sequence: MLAAAFADSN SSSMNVSFAH LHFAGGYLPS DSQDWRTIIP
 45 ALLVAVCLVG FVGNLCVIGI LLHNAWKGP SMHSLILNL
 SLADLSLLF SAPIRATAYS KSVWDLGWFW CKSSDWFIHT
 CMAAKSLTIV VVAKVCFMYA SDPAKQVSIH NYTIWSVLVA
 IWTVASLLPL PEWFSTIRH HEGVEMCLVD VPAVAEEFMS
 MFGKLYPLLA FGLPLFFASF YFWRAYDQCK KRGTKTQNLR
 50 NQIRSKQVTM MLLSIAIISA LLWLPEWWAW LWVWHLKAAG
 PAPPQGFIAL SQVLMFSISS ANPLIFLVMS EEFREGLKGV
 WKWMITKKPP TVSESQETPA GNSEGLPDKV PSPESPASIP
 EKEKPSSPSS GKKGTEKAEI PILPDVEQFW HERDTVPSVQ
 DNDPIPWEHE DQETGEGV

55

SEQ ID NO:19

189885

Cluster name: G protein-coupled receptor Ls189885

5 SequenceID: ENSMDNA178311

Sequence: GGGGCTTCGG AGGTGATCGG GCAGTGTCAAG TCTTCAGCCA
CTAAGCCGAG AAGATCTGGG AAGGAATCAG TCAGAGAGCC
TTGGGCCAGA GTTCCAGGGG CTCTGGGAGT GGGTGTCAAGA
GAGATTGACC AAACATTAGG AATTGACACC ATTCTCTGTC
10 ACCATCATGA AAGACTTCTT CAGTCTCATT ACGGAATTCA
CAAGTCTTCT TTAATGTCAG TAGGAAATTCA ACAAGTCGCA
GCTTGTACAG AGCTGAATGT TTATGTTGTT GCTGACACAG
TTGGATTAAAT TATCAAATCC AATTCAATCC TGGACTCAGT
CCAGCCTAAC TATTGCTCAA ATAAACACAT AGAGCTCAGA
15 ACACAAGTTG GTGGAGCTCG GAATCTGAGA GCAAACACTCAC
CCATGACCTC CAGCTACAAT CAAGAGAGCA GTAGCATGGAA
GAATGTGTCT GCATTGTCAC TGTTGACTGT GGAGAGTCCTC
ACGTCCATGT TTGACTATTG TGATGACTCT TTGGAGAGGG
TCAAGTCTGC TCTTGACATC TTTTCCATGA TCATCTACAC
20 AGTGACTTTTC TTCTTAGGCT TGGCTGGAA TGGCCTTGTC
ATTTGGGTAG TTGGATTCCA CATGTCCTGC ACAGTCAACA
CGTGTCTTCC TTCTGACCCCT CATCTCCATG GACCAACTGAC
TTGTGATCTC GTGGCCAATC TAGTCCCTGGA ACAATTGAC
ACCAGCAAAG GCAACTCTGG GGCCCTTGAG GACCTGGCTT
25 TTGGCAATTGTTT GTTTCTCTGT TCCCTACTTGT ATCTTCAGG
AAACTCGTGG TGGAAAGTGT CACCCCTCTT GTACAACCCAG
TATGATCTGC AGAATGAAAC TCAAGGAAGT CACCAACTTT
GGAAAGAGAT TATCATTCCA TGGCACCAAA CGCTGGTCAC
AACAGCCCCAC TTTTCTTTG GCTTCTTCT CCCTCTGGCT
30 ATCATCACTG GCTACTACAT CCTTGTAGCC TTGAAGTTAA
GAGAAAGGCA GCTGGTTAAG TTTAGCTGA

SEQ ID NO:20

189886

35 Cluster name: G protein-coupled receptor Ls189886

SequenceID: AI659965

Sequence: ACGTATTTTT TATTTTATCA CAACGTACAC GGATGAGACA
TTCCCCACTC AAGAAAAGTGT ATGTGAAGTCTGCCTTGAA
GAGAGTCAAA TGTCCAAAAC GTAGCCGGAA ATTGGAAGAT
40 GCAAGAAGCA TCAGGAGAGA AGAGGGTCTC TGGGGGACAG
CGACTGGGGA GGGCTTGAGG CAGGACTCCA CGCTTATTCC
TGTCTGAACC GCCGGAGTGT GGGGGGACGG TGGGGGCAGA
GGGAAAGGCC AGGGACTGTC GTCAGGAACA TCGCCTTGGC
AGGAAAGCAC GCATTCTATT AGGTTGGTGC ACAAAATCACG
45 GCAGAACACAGC AGTTTGACAC CAACCTAATG CTTACAAAAA
CACAAAATCA CCCACGTCAA AATGCTCCAT AAATGGCATC
AGACTTGGCC GGGCGCAGTG GCTCACGGCT GGGTAATGGT
CCACGCTCAC ACAGGCCATG AGGTAGACCC CCCCCTAGGT
50 TTGCCGAAAG CCAGCTGGAG CCCAGCACAT AACACACCCAC
CCTTCCGGT AAGGCCAGGT GGAACAGCAG TCAG

SEQ ID NO:21

LS_ID 189889

Cluster name: G protein-coupled receptor Ls 189889

SequenceID: ENSMDNA37702

Sequence: ATGCATGTGG GCAGGTATGA AGGACACCCA GACACAGGAG
CAGACAAACAT GCTGAGAGTG ATATGCTTTC CTTCATGAA
5 GGTGTCAAGG AGCCGGCAGC ACAGTGGATG TGCAGACCAT
GAAGGTGACC CCAAAATCTG CCTGGTGCAC AGCACAAAGTG
ATGGGGTCTG GGTGGCCAAT GAACATGAAG GGGCAGAGGA
AGCTGAGGGC CAAGGAGGAC AGCAGGAGAT AGCTGAGCTG
10 GCAGTTGTTG GCTCGGATGA TGGGAGTGTG GTGGTGTCAAG
ACGAAGATGC CTAA

SEQ ID NO: 22

189895

Cluster name: G protein-coupled receptor GPR61

15 SequenceID: AF317652

Sequence: ATGGAGTCCT CACCCATCCC CCAGTCATCA GGGAACTCTT
CCACTTTGGG GAGGGTCCCT CAAACCCCAG GTCCCTCTAC
TGCCAGTGGG GTCCCGGAGG TGGGGCTACG GGATGTTGCT
TCGGAATCTG TGGCCCTCTT CTTCATGCTC CTGCTGGACT
20 TGACTGCTGT GGCTGGCAAT GCGCCTGTGA TGGCCGTGAT
CGCCAAGACG CCTGCCCTCC GAAAATTGTT CTTCGTCTTC
CACCTCTGCC TGGTGGACCT GCTGGCTGCC CTGACCCCTCA
TGCCCCCTGGC CATGCTCTCC AGCCCTGCC TCTTTGACCA
CGCCCTCTTT GGGGAGGTGG CCTGCCGCCT CTACTTGTGTT
25 CTGAGCGTGT GCTTTGTCAG CCTGGCCATC CTCTCGGTGT
CAGCCATCAA TGTGGAGCGC TACTATTACG TAGTCCACCC
CATGCGCTAC GAGGTGCGCA TGACGCTGGG GCTGGTGGCC
TCTGTGCTGG TGGGTGTGTG GGTGAAGGCC TTGGCCATGG
CTTCTGTGCC AGTGTGTTGGAG AGGGTCTCCT GGGAGGAAGG
30 AGCTCCCAGT GTCCCCCAGC ACTGTTCACT CCAGTGGAGC
CACAGTGCCT ACTGCCAGCT TTTTGTGGTG GTCTTGTGCT
TCCTTITACCT TCTGTGCTCC CTGCTCCTCA TACTTCTGGT
CTACTGCAGC ATGTTCCGAG TGGCCCGCGT GGCTGCCATG
CCAGACGGGC CGCTGCCAC GTGGATGGAG ACACCCCGGC
35 AACGCTCCGA ATCTCTCAGC AGCGCTCCA CGATGGTCAC
CAGCTCGGGG GCCCCCCAGA CCACCCACA CGGGACGTTT
GGGGGAGGGGA AAGCAGCAGT GGTTCTCCTG GCTGTGGGGG
GACAGTTCT GCTCTGTGG TTGCCCTACT TCTCTTCCA
CCTCTATGTT GCCCTGAGTG CTCAGCCCCAT TTCAACTGGG
40 CAGGTGGAGA GTGTGGTCAC CTGGATTGGC TACTTTGCT
TCACCTTCCAA CCCTTCTTC TATGGATGTC TCAACCGGC
GATCCGGGGG GAGCTCAGCA AGCAGTTGT CTGCTTCTTC
AAGCCAGCTC CAGAGGAGGA GCTGAGGCTG CCTAGCCGGG
AGGGCTCCAT TGAGGAGAAC TTCCTGCAGT TCCTTCAGGG
45 GACTGGCTGT CCTCTGAGT CCTGGGTTTC CCGACCCCTA
CCCAGCCCCA AGCAGGAGGC ACCTGCTGTT GACTTTCGAA
TCCAGGCCAG ATAG

SEQ ID NO: 23

50 189897

Cluster name: G protein-coupled receptor GPR73

SequenceID: AR070166

Sequence: AGCCGCAGAG CGCACAGAAA GGAGGCAGCG AGACAGACAT
CACCATGGCA GCCCAGAATG GAAACACCAG TTTCACACCC

AACTTTAAC CACCCCAAGA CCATGCCCTCC TCCCCTCCTC
TTAACCTCAG TTATGGTGTATGACCTCC CTATGGATGA
GGATGAGGAC ATGACCAAGA CCCGGACCTT CTTCGCAGCC
AAGATCGTCA TTGGCATTGC ACTGGCAGGC ATCATGCTGG
5 TCTCGGGCAT CGGTAACCTT GTCTTATCG CTGCCCTCAC
CCGCTATAAG AAGTTGCGCA ACCTCACCAA TCTGCTCATT
GCCAACCTGG CCATCTCCGA CTTCTGGTG GCCATCATCT
GCTGCCCTC CGAGATGGAC TACTACGTGG TACGGCAGCT
CTCCTGGGAG CATGGCCACG TGCTCTGTGC CTCCGTCAAC
10 TACCTGCGCA CCGTCTCCCT CTACGTCTCC ACCAATGCCT
TGCTGGCCAT TGCCATTGAC AGATATCTCG CCATCGTCA
CCCCTGAAA CCACGGATGA ATTATCAAAC GGCCTCCTTC
CTGATCGCT TGGTCTGGAT GGTGTCATT CTCAATTGCCA
TCCCATCGGC TTACTTTGCA ACAGAAACCG TCCTCTTAT
15 TGTCAAGAGC CAGGAGAAGA TCTTCTGTGG CCAGATCTGG
CCTGTGGATC AGCAGCTCTA CTACAAGTCC TACTTCCTCT
TCATCTTGG TGTGGAGTTG TGCCCCCTG TGGTCACCAT
GACCCTGTGC TATGCCAGGA TCTCCGGGA GCTCTGGTTC
AAGGCAGTCC CTGGGTTCCA GACGGAGCAG ATTGCAAGC
20 GGCTGCGCTG CCGCAGGAAG ACGGTCTGG TGCTCATGTG
CATTCTCACG GCCTATGTGC TGTGCTGGGC ACCCTTCTAC
GGTTTCACCA TCGTTCGTGA CTTCTCCCC ACTGTGTTCG
TGAAGGAAAA GCACTACCTC ACTGCCTTCT ACGTGGTCGA
GTGCATCGCC ATGAGCAACA GCATGATCAA CACCGTGTGC
25 TTCGTGACGG TCAAGAACAA CACCATGAAG TACTICAAGA
AGATGATGCT GCTGCACTGG CGTCCCTCCC AGGGGGGGAG
CAAGTCCAGT GCTGACCTTG ACCTCAGAAC CAACGGGGTG
CCCACCACAG AAGAAGTGGGA CTGTATCAGG CTGAAGTGAC
CCACTGGTGT CACACAATTG AAAACCCAG TCCAGTACTC
30 AGAGCATCAC CCACCATCAA CCAAGTTCAT AGGCTGCATG
GGAAATGACA TCTGTGTTCA TGCCCTCCCC GTGCCCTCAA
GAAGCCGAAT GCTGCAAAGT CGTAACATAC AATGAGACTA
GACATGAACC AAATCAGCTG ACATTTACTG ATATCCGCTC
GACACCTACT GTGTCACAA TCCCCACAAG GAGATTAGAC
35 ACAAGGAGCA GCAACTGACA TGGACTGAAC ATGTACTGTG
TGCAAACAC ACCAATGAGA TTAGACGGGG ACAGCAGGAG
CTGACATTCA CTCTTCACCT ACTGTAATCA AAAACACTTG
ATTGATTAC AATCAAAAAC ATATAAAAAA CATAACAAAG
TAGCAGAACG TATTGGAGIT TCCAAGCTAT CTCCAGATAT
40 ATAGATAGTT CACCCCTCCAT CTTCCCTAAT TCTGTATCTT
ACCAAGTGCAG GAATATCAAAGGCTATAGG CCAGGCATGA
TGGCTCATGC CTGTAATCCC AGCACTGGG GAGGCTGAGG
CACGTGGATC ACTTGAGGTC AGGAGTCAA CCCAGGCTGG
CCAACATGGT GAAACCCGTCTCTACTAAA AATACAAAT
45 TAGCTAGGCG TGGTGGCGGG CGCCTGTAAT CCCAGTTACT
CAGGAGGCTG AAGCAGGAGA ATAGCTGAA CCTGGGAGTT
GGAGTTGCA GTGAGCTGAG ATTGCTCCAC TGCACTCCAG
CCTGAGTGAC AGAGTGAGAC TCTGTCTCAG GAAAAAAACA
AACAAACAAA CAACAAACACA ACAACAACAA CAACAACAC
50 CAACGGCTAT AGAAGAAGAC TCTTCGACAC AATGGAAATG
TAACGATAAG TTTGTCAGTG CGTGGTTAC AGCATCATGG
GAGGTGCGTT ACAGCCATCA TACTGAACCTT TCCCACCCAC
CTCCTACTGC CTCCCAGGGC ATTCTCTAGG ATTTGGCTT
CAAGAAAAAA AAAATTCTTA TAGTCAGCCC AGCCTTATGT
55 GTTATCCAC AATGGGTGAA TTCAAAGGA AAGAACCTAA
AAATCACTTT CCCACTGATG CTTGAAAGCT TATCATTTTA
TTTGGGTGGA GATGGGTAAT CCTGAGGTGT CAATTTTGCG
CTCCTCAGTG CAAAGGATT CAGTGGCTCT GGGGTCAAGGG
GGAAAGAGGA CAGAGAAAAA AGTGGAGGTT GCCACTGGCA
60 ATGAACATAA TCTCTGTGGG CATTTGCTA AGGACTGGAC

CACTTCTAG AACACTCCCT CTTTACAAA AGGAACCTA
 CCTAGAACCTC AAAGACCTGG GTTCAGGTCC TAACTCTAAG
 ACTCAAGTCC TAAATTCATG ATGTTTCTC TCTGTGTCTC
 AGTTTGCTT TAATGAAATG GCGATGATGA AAATATCTGC
 5 TCTCATACC TTGCAAGACT GTTGGGAGAG CCCATTGAGG
 CCATGGTTTG TGAATGTGCT TTCAACTGT GCACACGATA
 AGAATGGAGA AGTGATATTG AACAGTTAT TTGGAGGGAG
 TTTATTGGA AACCCCATCC ACTGTGATT ATTAGAGAAA
 TACCCACACT TTTCATCCC TGTTCTTGG ATGAAAGACT
 10 CCTGAAGACT TCACAGTGTA CTTGTCTAC AGTGGGCCAA
 AAAGGGATCC CTGTTCTGG TTATAATCTG GGAAATTAA
 CCTCAGATTC TCAGTGACCC CAAGACTCTC AGCATCCCTG
 CGGTCTTAGA AGTGTGACA GTCTCCCTG CATGTTGCAA
 AATAGCACCC TAGTGCTGCA TAAATATCAC TTCTGAATCT
 15 GTTGTATTAA TTATACATT GTGGTAACGT TAGGTACACG
 TCTTCATTTCTC TTCTTGATTCTTGTGTTGATGT GGTAGCTATG
 CAAATGGTAC CTGGTTGGG ACTGACCCAT CCATATTGAA
 CCAATTCTA ATTGTTATA GACAAGGAAT TAATTGTTG
 CTTGTTGAT TGTTCTATT ATTGTTGAT TTGTTCTCT
 20 GACTGAAGTT TCAACCAATG TTCTTCTA TCACCACCCA
 GCAGACTCAC CTTCAGGCCA ATCATTGTAC TCTCAGAAAA
 TGCAGGCCGG CATGGTGGCT CACATCTGTA ATCCCAGCAC
 TTCGGGAGGC CAAGATGGGC AGATCACCTG AGGTCAAGGAG
 TTCAAGACCA GCCTGGCCAA CATGGCAAAA CCCCCATCTCT
 25 AGAAAAATAC AGAAATTAGC TGGCGTGGTG GCACATGCC
 GTGGTCCCGAG CTCCTCAGGA GGCTGAGGCA TGAGAATTGC
 TTGAACCCCA GAGGCAGAGG TTGCAGTGAA TTGAGATCGC
 ACCACTGCAC TCCAGCCTGG GTGATAGAGC AAGATTCCAT
 CTCAAAAGGA AAATAAAAGA AAATGCAAAC ACACATAAT
 30 ATTAGCCTAA GCAAAACTGT TAATTCTGAT TTACAAAAAT
 TCTTACTTGC TTGGCTTGA AATGCATTGT GTAATAATGC
 ATTCAAAGC CAAGCAAGTA ACAATTITAG GITATGTACA

SEQ ID NO: 24

35 189900

Cluster name: Sphingosine 1-phosphate receptor Edg-8

SequenceID: AF317676

Sequence: ATGGAGTCGG GGCTGCTGCG GCCGGGCCG GTGAGCGAGG
 40 TCATCGTCCT GCATTACAAC TACACCGGCA AGCTCCCGGG
 TGCGCGCTAC CAGCCGGGTG CCGGCCTGCG CGCCGACGCC
 GTGGTGTGCC TGGCGGTGTG CGCCTTCATC GTGCTAGAGA
 ATCTAGCCGT GTTGTGGTG CTCGGACGCC ACCCGCGCT
 CCACGCTCCC ATGTTCTGC TCCGGGCAG CCTCACGTG
 TCGGATCTGC TGGCAGGCC CGCCTACGCC GCCAACATCC
 45 TACTGTCGGG GCCGCTCACG CTGAAACTGT CCCCCCGCGCT
 CTGGTTCGCA CGGGAGGGAG GCGTCTTCGT GGCACACTCACT
 GCGTCCTGTC TGAGCCTCT GGCCATCGCG CTGGAGCGCA
 GCCTCACCAT GGCGCGCAGG GGGCCCGCGC CGCTCTCCAG
 TCGGGGGCGC ACGCTGGCGA TGGCAGCCGC GGCGCTGGGC
 50 GTGTCGCTGC TCCCTGGGCT CCTGCCAGCG CTGGGCTGG
 ATTGCGCTGGG TCGCCTGGAC GCTTGCTCCA CTGTCCTGCC
 GCTCTACGCC AAGGCCTACG TGCTCTCTG CGTGCTCGCC
 TTCGTGGGCA TCCGGGCCGC GATCTGTGCA CTCTACGCC
 GCATCTACTG CCAGGTACGC GCCAACCGCGC GGCGCTGCC
 55 GGCACGGCCC GGGACTCGGG GGACCCACCTC GACCCGGCG
 CGTCGCAAGC CGCGCTCGCT GGCGCTGCG CGCACGCTCA
 GCGTGGTGCT CCTGGCCTTT GTGGCATGTT GGGGCCCCCT
 CTTCCTGCTG CTGTTGCTCG ACGTGGCGTG CCCGGCGCGC

ACCTGTCCTG TACTCCTGCA GGCGATCCC TTCCCTGGAC
TGGCCATGGC CAACTCACTT CTGAACCCCA TCATCTACAC.
GCTCACCAAC CGCGACCTGC GCCACCGCGCT CCTGCGCCTG
GTCTGCTGCG GACGCCACTC CTGCGGCAGA GACCCGAGTG
5 GCTCCCAGCA GTCGGCGAGC GCGGCTGAGG CTTCCGGGGG
CCTGCGCCCG TGCCTGCCCC CGGGCCTGA TGGGAGCTTC
AGCGGCTCGG AGCGCTCATC GCCCCAGCGC GACGGGCTGG
ACACCAAGCGG CTCCACAGGC AGCCCCGGTG CACCCACAGC
CGCCCGGACT CTGGTATCAG AACCGGCTGC AGACTGA

10

SEQ ID NO: 25

189901

Cluster name: G protein-coupled receptor Ls189901

SequenceID: E31720

15 Sequence: GACTATCCTC CCACTTCAGG GTTTCTCTGG GCTTCCATCT
TGCCCCTGCT GAGCCCTGCT TCCTCCTCTA CCAGCAGCAC
AACCCCCAGG CTGGGCTCAG AGACCTCATG TGGTGGGATC
ACTCAGTACC CCGAGGGCGGA GGGAAAGGAGG GAGGGCTGCA
GGGTTCCCCT TGGCCTGCAA ACAGGAACAC AGGGTGTTC
20 TCAGTGGCTG CGAGAAATGCT GATGAAAACC CCAGGGATGTT
GTGTCAACCGT GGTGGCCAGC TGATAGTGCC AATCATCCCA
CTTGTCCCTG AGCACTCTG CAGGGGTAGA AGACTCCAGA
ACCTTCTCTC AGGCCATGG CCCAAGCAGC CCATGGAACCT
TCATAACCTG AGCTCTCCAT CTCCCTCTCT CTCCCTCCTCT
25 GTTCTCCCTC CCTCCCTCTC TCCCTCACCC TCCTCTGCTC
CCTCTGCCTT TACCACTGTG GGGGGGTCT CTGGAGGGCC
CTGCCACCCCC ACCTCTTCCT CGCTGGTGTG TGCCCTTCCTG
GCACCAAATCC TGGCCCTGGA GTTTGTCTG GCCCTGGTGG
30 GGAACAGTTT GGCCCTCTTC ATCTTCTGCA TCCACACGCG
GCCCTGGACC TCCAACACGG TGTTCTGTG CAGCCTGGTG
GCCGCTGACT TCCTCTTGAT CAGCAACCTG CCCCTCCGCG
TGGACTACTA CCTCCTCCAT GAGACCTGGC GCTTTGGGGC
TGCTGCCTGC AAAGTCAACC TCTTCATGCT GTCCACCAAC
CGCACGGCCA GCGTTGTCTT CCTCACAGCC ATCGCACTCA
35 ACCGCTACCT GAAGGTGGTG CAGCCCCACC ACGTGCTGAG
CCGTGCTTCC GTGGGGGCAG CTGCCCCGGGT GGCCGGGGGA
CTCTGGGTGG GCATCCTGCT CCTCAACGGG CACCTGCTCC
TGAGCACCTT CTCCGGCCCC TCCTGCCTCA GCTACAGGGT
GGGCACGAAG CCCTCGGCCT CGCTCCGCTG GCACCCAGGCA
40 CTGTACCTGC TGGAGTTCTT CCTGCCACTG GCGCTCATCC
TCTTTGCTAT TGTGAGCATT GGGCTCACCA TCCGGAACCG
TGGTCTGGGC GGGCAGGCAG GCCCCAGAG GGCCATGCGT
GTGCTGGCCA TGGTGGTGGC CGTCTACACC ATCTGCTCT
TGCCCAGCAT CATCTTGGC ATGGCTTCA TGGTGGCTTT
45 CTGGCTGTCC GCCTGCCGCT CCCTGGACCT CTGCACACAG
CTCTTCCATG GCTCCCTGGC CTTCACCTAC CTCAACAGTG
TCCTGGACCC CGTGCTCTAC TGCTTCTCTA GCCCCAACTT
CCTCCACCAAG AGCCGGGCCT TGCTGGGCCT CACCGGGGGC
CGGCAGGGCC CAGTGAGCGA CGAGAGCTCC TACCAACCC
50 CCAGGCAGTG GCGCTACCGG GAGGCCTCTA GGAAGGCGGA
GCCATAGGG AAGCTGAAAG TGCAGGGCGA GGCTCTCTG
GAAAAGGAAG GCTCCTCCCA GGGCTGAGGG CCAGCTGCAG
GGCTGCAGCG CTGTGGGGGT AAGGGCTGCC GCGCTCTGGC
CTGGAGGGAC AAGGCCAGCA CACGGTGCCT CAAC

55

SEQ ID NO: 26

190188

Cluster name: G protein-coupled receptor LGR6

SequenceID: AB049405

Sequence: GCCACTGCCA GGAGGGACGGC ATCATGCTGT CTGCCGACTG
 5 CTCTGAGCTC GGGCTGTCCG CCGTTCGGG GGACCTGGAC
 CCCCTGACGG CTTACCTGGA CCTCAGCATG AACAAACCTCA
 CAGAGCTTC A GCCTGGCCTC TTCCACCACC TGCGCTTCTT
 GGAGGAGCTG CGTCTCTCTG GGAACCATCT CTCACACATC
 CCAGGACAAG CATCTCTGG TCTCTACAGC CTGAAAATCC
 10 TGATGCTGCA GAACAATCAG CTGGGAGGAA TCCCCGCAGA
 GGCCTGTGG GAGCTGCCGA GCCTGCAGTC GCTGCGCTA
 GATGCCAACC TCATCTCCCT GGTCCCAGG AGGAGCTTG
 AGGGGCTGTC CTCCCTCCGC CACCTCTGGC TGGACGACAA
 TGCACTCACG GAGATCCCTG TCAGGGCCCT CAACAACCTC
 15 CCTGCCCTGC AGGCCATGAC CCTGGCCCTC AACCGCATCA
 GCCACATCCC CGACTACGCG TTCCAGAAC TCAACCAGCCT
 TGTGGTGTG CATTTCATA ACAACCGCAT CCAGCAGTC
 GGGACCCACA GCTTCGAGGG GCTGCACAAT CTGGAGACAC
 TAGACCTGAA TTATAACAAG CTGAGGGAGT TCCCTGTGGC
 20 CATCCGGACC CTGGGCAGAC TGCAGGAACG GGGGTTCCAT
 ACAACAAACA TCAAGGCCAT CCCAGAAAAG GCCTTCATGG
 GGAACCCCTC GCTACAGACG ATACACTTTT ATGATAACCC
 AATCCAGTTT GTGGGAAGAT CGGCATTCCA GTACCTGCCT
 AAACCTCCACA CACTATCTCT GAATGGTGCA ATGGACATCC
 AGGAGTTTC AGATCTCAA GGCACCACCA GCCTGGAGAT
 25 CCTGACCCCTG ACCCGCGCAG GCATCCGGCT GCTCCCATCG
 GGGATGTGCC AACAGCTGCC CAGGCTCCGA GTCCCTGGAAC
 TGTCTCACAA TCAAATTGAG GAGCTGCCA GCCTGCACAG
 GTGTCAGAAA TTGGAGGAAA TCGGCCTCCA ACACAACCGC
 ATCTGGAAA TTGGAGGCTGA CACCTTCAGC CAGCTGAGCT
 30 CCCTGCAAGC CCTGGATCTT AGCTGGAACG CCATCCGGTC
 CATCCACCCCT GAGGCCCTCT CCACCCCTGCA CTCCCTGGTC
 AAGCTGGACC TGACAGACAA CCAGCTGACC ACACTGCC
 TGGCTGGACT TGGGGGCTTG ATGCATCTGA AGCTCAAAGG
 35 GAACCTTGCT CTCTCCCAGG CCTCTCCAA GGACAGTTTC
 CCAAAACTGA GGATCCTGGA GGTGCCTTAT GCCTACCAAGT
 GCTGTCCCTA TGGGATGTGT GCCAGCTCT TCAAGGCCTC
 TGGGCAGTGG GAGGCTGAAG ACCTTCACCT TGATGATGAG
 GAGTCTTCAA AAAGGCCCCCT GGGCCTCCTT GCCAGACAAG
 CAGAGAACCA CTATGACCAG GACCTGGATG AGCTCCAGCT
 40 GGAGATGGAG GACTCAAAGC CACACCCAG TGTCCAGTGT
 AGCCCTACTC CAGGCCCTT CAAGCCCTGT GAGTACCTCT
 TTGAAAGCTG GGGCATCCGC CTGGCCGTGT GGGCCATCGT
 GTGCTCTCC GTGCTCTGCA ATGGACTGGT GCTGCTGACC
 GTGTTCGCTG GCAGGCCCTGC CCCCCCTGCC CCGGTCAAGT
 45 TTGTGGTAGG TGCAGATTGCA GGCGCAACA CCTTGACTGG
 CATTTCCTGT GGCCTCTAG CCTCAGTCGA TGCCCTGACC
 TTTGGTCAGT TCTCTGAGTA CGGAGCCCGC TGGGAGACGG
 GGCTAGGCTG CCGGGCCACT GGCTCCTGG CAGTACTTGG
 GTCGGAGGCA TCGGTGCTGC TGCTCACTCT GGCCGAGTG
 50 CAGTGCAGCG TCTCCGTCTC CTGTGTCCGG GCCTATGGGA
 AGTCCCCCTC CCTGGGCAGC GTTCGAGCAG GGGTCCCTAGG
 CTGCCCTGGCA CTGGCAGGGC TGGCCGCCGC ACTGCCCTG
 GCCTCAGTGG GAGAATACGG GGCCCTCCCCA CTCTGCCTGC
 CCTACGCGCC ACCTGAGGGT CAGCCAGCAG CCCTGGGCTT
 55 CACCGTGGCC CTGGTGATGA TGAACCTCTT CTGTTTCTG
 GTCGTGGCCG GTGCTACAT CAAACTGTAC TGTGACCTGC
 CGCGGGGGCGA CTTTGAGGCC GTGTGGGACT GCGCCATGGT
 GAGGCACGTG GCCTGGCTCA TCTTCAGCA CGGGCTCCTC
 TACTGTCCCCG TGGCCCTCCT CAGCTTGTGCC TCCATGCTGG

GCCTCTTCCC TGTACGCCCGAGGGCGTCA AGTCTGTCCT
GCTGGTGGTG CTGCCCTGCACTGCCT CAACCCACTG
CTGTACCTGC TCTTCAACCC CCACTTCCGG GATGACCTTC
GGCGGCTTCG GCCCGCGCA GGGGACTCAG GGCCCCTAGC
5 CTATGCTGCG GCCGGGGAGC TGGAGAAAGAG CTCCGTGAT
TCTACCCAGG CCCTGGTAGC TTTCTGTGAT GTGGATCTCA
TTCTGGAAGC TTCTGAAGCT GGGCGCCCC CTGGGCTGGA
GACCTATGGC TTCCCCTCAG TGACCCCTCAT CTCCTGTAG
CAGCCAGGGG CCCCCAGGCT GGAGGGCAGC CATTGTGAG
10 AGCCAGAGGG GAACCACTTT GGGAAACCCCC AACCCCTCAT
GGATGGAGAA CTGCTGCTGA GGGCAGAGGG ATCTACGCCA
GCAGGGTGGAG GCTTGTCAAGG GGGTGGCGGC TTTCAGCCCT
CTGGCTTGGC CTTTGTCTCA CACGTGAAA TATCCCTCCC
CATTCTCTC TTCCCCTCTC TTCCCCTTC TCTCTCCCC
15 TCGGTGAATG ATGGCTGCTT CTAAACAAA TACAACCAA
ACTCAGCAGT GTGATCTATA GCAGGATGGC CCAGTACCTG
GCTCCACTGA TCACCCCTCTC CCTGTGACCA TCACCAACGG
GTGCCTCTTG GCCTGGCTT CCCTTGGCCT TCCTCAGCTT

20 SEQ ID NO:27

190411

Cluster name: G protein-coupled receptor Ls190411

SequenceID: AF305409

Sequence: CCACAAGGAG TAGTTGGGAG ATACAGGGGC ATGGCCACCA
25 CAAGCAGAAC AATTTCGGG ATATTTGTA GAAGATGGGG
TTTGCCACA TTGCCAGGC TGGTCTCGAA CTGGGTGGGA
TCAAAACGATC CAACCGCGTT GGCCTCCAGA GTGTTGGGAT
TACAGGTGTG AGCCACCAAG CATGGAATAG GCTTCTTAA
ACATTGAATA GTATTCCTT GGTAGATGAA GGAGGGATGAG
30 ATAGCACGAG AGGGCAAAGA TGCAGCCAAG TAACCCAGTG
CTGGAGCCA CGATGGAGAA GATCTCACGG CCACCTCTGGC
CTTGCCTGG GTGCTTAGT AACTCGGGAG GAAGGCCACC
CAGACACTGC AGGACACCAAG CATGCTGAAG GTCAAGGA
TGACTTATTG AAGGTGTAG GCAGGTTCT TGCCAGAAAG
35 GCTACAGCAA GGGACCTAA ACCAAGAAG CCCAAGTAGC
CCAAGACAGA GTAGAAGGCA GTGACGGAGC CCTCATTACA
CTGGATAATG ATGTAGGCCAG GCATGAACGT AGGGTCTTG
TTACGAAGG GAGGCTCTGT CCCCCAGCCAG ATTCCACAGA GGGTC

40

SEQ ID NO:28

190414

Cluster name: G protein-coupled receptor Ls190414

SequenceID: AX080495

45 Sequence: GCCTGCAACC TGTCYCACGC CCTCTGGCTG TTGCCATGAC
GTCCACCTGC ACCAACAGCA CGCGCGAGAG TAACAGCAGC
CACACGTGCA TGCCCTCTC CAAAATGCC ATCAGCCTGG
CCCACGGCAT CATCCGCTCA ACCGTGCTGG TTATCTTCT
CGCCGCCTCT TTCTGCGCA ACATAGTGT GGCCTAGTG
50 TTGCAGCGCA AGCCGAGCT GCTGCAGGTG ACCAACCGTT
TTATCTTAA CCTCCCTCGTC ACCGACCTGC TGCAAGATTTC
GCTCGTGGCC CCCTGGGTGG TGGCCACCTC TGTGCCTCTC
TTCTGGCCCC TCAACAGCCA CTTCTGCACG GCCCTGGTTA
GCCTCACCCA CCTGTTGCC TTGCGCAGCG TCAACACCAT
55 TGTCTTGGTG TCAGTGGATC GCTACTTGTGTC CATCATCCAC

CCTCTCTCCT ACCCGTCCAA GATGACCCAG CGCCGCGGTT
 ACCTGCTCCT CTATGGCACC TGGATTGTGG CCATCCTGCA
 GAGCACTCCT CCACTCTACG GCTGGGCCA GGCTGCCTT
 5 GATGAGCGCA ATGCTCTCG CTCCATGATC TGGGGGCCA
 GCCCCAGCTA CACTATTCTC ACGTGGTGT CCTTCATCGT
 CATTCCACTG ATTGTATGA TTGCCTGCTA CTCCGTGGTG
 TTCTGTGCAG CCCGGAGGCA GCATGCTCTG CTGTACAATG
 TCAAGAGACA CAGCTGGAA GTGCGAGTC AGGACTGTGT
 GGAGAATGAG GATGAAGAGG GAGCAGAGAA GAAGGAGGAG
 10 TTCCAGGATG AGAGTGAGTT TCGCCGCCAG CATGAAGGTG
 AGGTCAAGGC CAAGGAGGGC AGAACATGGAG CCAAGGACGG
 CAGCCTGAAG GCCAAGGAAG GAAGCACGGG GACCAGTGAG
 AGTAGTGTAG AGGCCAGGGG CACCGAGGAG GTCAAGAGAGA
 GCAGCACGGT GGCCAGCGAC GGCAGCATGG AGGGTAAGGA
 15 AGGCAGCACC AAAGTTGAGG AGAACACGCAT GAAGGCAGAC
 AAGGGTCGCA CAGAGGTCAA CCAGTGCAGC ATTGACTTGG
 GTGAAGATGG CATGGAGTT GGTGAAGACG ACATCAATT
 CAGTGAGGAT GACGTCGAGG CAGTGAACAT CCCGGAGAGC
 20 CTCCCACCCA GTCGTCGTA CAGCAACAGC AACCCCTCCTC
 TGCCCAGGTG CTACCACTGC AAAGCTGCTA AAGTGATCTT
 CATCATCATT TTCTCCTATG TGCTATCCCT GGGGGCCCTAC
 TGCTTTTAG CAGTCCTGGC CGTGTGGGTG GATGTCGAAA
 CCCAGGTACC CCAGTGGGTG ATCACCATAA TCATCTGGCT
 TITCTTCCTG CAGTGCTGCA TCCACCCCTA TGTCTATGGC
 25 TACATGCACA AGACCATTAA GAAGGAAATC CAGGACATGC
 TGAAGAAGTT CTTCTGCAAG GAAAAGCCCC CGAAAGAAGA
 TAGCCACCCA GACCTGCCCG GAACAGAGGG TGGGACTGAA
 GGCAAGATTG TCCCTTCCTA CGATTCTGCT ACTTTTCCTT
 GAAGTTAGTT CTAAGGCAAAC CCTTGAAAAT CAGTCCTTCA
 30 GCCACAGCTA TTTAGAGCTT TAAAACCTACC AGGTTCAATC
 ACTGGTTATG CTTCTGTG

SEQ ID NO:29

190418

35 Cluster name: G protein-coupled receptor EX33 (GPR84)

SequenceID: NM_020370

Sequence: TAACTGTCCA CCAGAAAGGA CTGCTCTTG GGTGAGTTGA
 ACTTCTTCCA TTATAGAAAG AATTGAAGGC TGAGAAACTC
 40 AGCCTCTATC ATGTGGAACA GCTCTGACGC CAACCTCTCC
 TGCTACCATG AGTCTGTGCT GGGCTATCGT TATGTTGCAG
 TTAGCTGGGG GGTGGTGGTG GCTGTGACAG GCACCGTGGG
 CAATGTGCTC ACCCTACTGG CCTTGGCCAT CCAGCCCAAG
 CTCCGTACCC GATTCAACCT GTCATAGCC AACCTCACAC
 TGGCTGATCT CCTCTACTGTC ACGCTCCCTC AGCCCTTCTC
 45 TGTGGACACC TACCTCCACC TGCACCTGGCG CACCGGTGCC
 ACCTTCTGCA GGGTATTGG GCTCCTCCIT TTTGCCTCCA
 ATTCTGTCTC CATCTGACC CTCTGCCCTCA TCGCACTGGG
 ACGCTACCTC CTCATGCCC ACCCTAAGCT TTTCCCCAA
 50 GTTTTCAGTG CCAAGGGGAT AGTGCTGGCA CTGGTGAGCA
 CCTGGGTGT GGGCGTGGCC AGCTTGCTC CCCTCTGGCC
 TATTTATATC CTGGTACCTG TAGTCTGCAC CTGCAGCTT
 GACCGCATCC GAGGCCGGCC TTACACCACC ATCCTCATGG
 GCATCTACTT TGTGCTTGGG CTCAGCAGTG TTGGCATCTT
 CTATTGCCTC ATCCACCGCC AGGTCAAACG AGCAGCACAG
 55 GCACTGGACC AATACAAGIT GCGACAGGCA AGCATCCACT
 CCAACCATGT GGCCAGGACT GATGAGGCCA TGCCCTGGTCG
 TTCCAGGAG CTGGACAGCA GGTTAGCATC AGGAGGACCC
 AGTGAGGGGA TTTCATCTGA GCCAGTCAGT GCTGCCACCA

CCCAGACCT GGAAGGGGAC TCATCAGAAG TGGGAGACCA
 GATCAACAGC AAGAGAGCTA AGCAGATGGC AGAGAAAAGC
 CCTCCAGAAG CATCTGCCAA AGCCCAGCCA ATTAAAGGAG
 CCAGAAGAGC TCCGGATTCT TCATCGGAAT TTGGGAAGGT
 5 GACTCGAATG TGTTTGCTG TGTTCTCTG CTTGCCCTG
 AGCTACATCC CCTTCTGCT GCTAACATT CTGGATGCCA
 GAGTCCAGGC TCCCCGGGTG GTCCACATGC TTGCTGCCAA
 CCTCACCTGG CTCAATGGTT GCATCAACCC TGTGCTCTAT
 GCAGCCATGA ACCGCCAATT CCGCCAAGCA TATGGCTCCA
 10 TTTAAAAAAG AGGGCCCCGG AGTTTCCATA GGCTCCATTA
 GAACGTGAC CCTAGTCACC AGAATTCAAG ACTGTCTCCT
 CCAGGACCAA AGTGGCCAGG TAATAGGAGA ATAGGTGAAA
 TAACACATGT GGGCATTTTC ACAACAATCT CTCCCCAGCC
 15 TCCCAAAATCA AGTCTCTCCA TCACTTGATC AATGTTTCAG
 CCCTAGACTG CCCAAGGAGT ATTATTAATT ATTAATAAT
 GAATTCTGTG CTTTAAAAA AAAAAAAATA AAAAAAGAAA
 AAAAAAAAAA AAAAAAAAAA AAAAAA

SEQ ID NO:30

20 190419

Cluster name: G protein-coupled receptor Ls190419

SequenceID: AJ303165

Sequence: CTTGCTTCA GAGCTAAACC AGTTTTCTT CTCTCACAG
 CAAATATCTT GACAGTGATC ATCCTCTCCC AGCTGGTGGC
 25 AAGAAGACAG AAGTCCTCCT ACAACTATCT CTTGGCACTC
 GCTGCTGCCG ACATCTGGT CCTCTTTTC ATAGTGTGG
 TGGACTTCCT GTTGAAGAT TTCATCTTGA ACATGCAGAT
 GCCTCAAGTC CCCGACAAGA TCATAGAAGT GCTGGAATT
 TCATCCATCC ACACCTCCAT ATGGATTACT GTACCGTTAA
 30 CCATTGACAG GTATATCGCT GTCTGCCACC CGCTCAAGTA
 CCACACGGTC TCATACCCAG CCCGCACCCG GAAAGTCATT
 GTAAGTGTTC ACATCACCTG CTTCTGACC AGCATCCCCT
 ATTACTGGTG GCCAACATC TGGACTGAAG ACTACATCAG
 CACCTCTGTG CATCACGTCC TCATCTGGAT CCACGTCTTC
 35 ACCGTCTACC TGGTCCCTG CTCCATCTTC TTCATCTTGA
 ACTCAATCAT TGTGTACAAG CTCAGGAGGA AGAGCAATT
 TCGTCTCCGT GGCTACTCCA CGGGGAAGAC CACCGCCATC
 TTGTTCACCA TTACCTCCAT CTTGCCACA CTTGGGCC
 CCCGCATCAT CATGATTCTT TACCACCTCT ATGGGGCGCC
 40 CATCCAGAAC CGCTGGCTGG TGACACATCAT GTCCGACATT
 GCCAACATGC TAGCCCTCT GAACACAGCC ATCAACTTCT
 TCCTCTACTG TTTCATCAGC AAGCGGTCTC GCACC

45 SEQ ID NO:31

190427

Cluster name: Cysteinyl leukotriene CysLT2 receptor

SequenceID: NM_020377

Sequence: AAGTTCTCTA AGTTGAAGC GTCAGCTTCA ACCAAACAAA
 50 TTAATGGCTA TTCTACATTC AAAAATCAGG AAATTTAAAT
 TTATTATGAA ATGTAATGCA GCATGTAGTA AAGACTTAAAC
 CAGTGTGTTA AAACTCAACT TTCAAAGAAA AGATAGTATT
 GCTCCCTGTT TCATTAACAC CTAGAGAGAT GTAATCAGTA
 AGCAAGAAGG AAAAAGGGAA ATTCAACAAAG TAACTTTTG
 55 TGTCTGTTTC TTTTAACCC AGCATGGAGA GAAAATTAT

GTCCITGCAA CCATCCATCT CCGTATCAGA AATGGAACCA
AATGGCACCT TCAGCAATAA CAACAGCAGG AACTGCACAA
TTGAAAACIT CAAGAGAGAA TTTITCCAA TTGTATATCT
5 GATAATATTT TTCTGGGGAG TCTTGGAAA TGGGTTGTCC
ATATATGTT TCCTGCAGCC TTATAAGAAG TCCACATCTG
TGAACTTTT CATGCTAAAT CTGGCCATT CAGATCTCCT
GTTCATATAAGC ACGCTTCCCT TCAGGGCTGA CTATTATCTT
AGAGGCTCCA ATTGGATATT TGGAGACCTG GCCTGCAGGA
TTATGTCITA TTCTTGTAT GTCAACATGT ACAGCAGTAT
10 TTATTTCTG ACCGTGCTGA GTGTTGCG TTTCTGGCA
ATGGTTCAACC CCTTTCGGCT TCTGCATGTC ACCAGCATCA
GGAGGTGCTG GATCCTCTGT GGGATCATAT GGATCCTTAT
CATGGCTTCC TCAATAATGC TCCTGGACAG TGGCTCTGAG
15 CAGAACGGCA GTGTCACATC ATGCTTAGAG CTGAATCTCT
ATAAAATTC TAAGCTGCAG ACCATGAACAT ATATTGCCCTT
GGTGGTGGGC TGCCCTGCTGC CATTTCAC ACTCAGCATC
TGTATCTGC TGATCATTG GGTCTGTT AAAAGTGGAGG
TCCCAGAACATC GGGGCTGCGG GTTTCTACA GGAAGGCAC
GACCACCATC ATCATCACCT TGATCATCTT CTTCTGTGT
20 TTCCTGCCCT ATCACACACT GAGGACCGTC CACTTGACGA
CATGGAAAGT GGGTTATGC AAAGACAGAC TGCATAAAGC
TTTGGTTATC ACACTGGCCT TGGCAGCAGC CAATGCCCTGC
TTCATCCTC TGCTCTATTAA CTTGCTGGG GAGAATTAA
AGGACAGACT AAAGTCTGCA CTCAGAAAAG GCCATCCACA
25 GAAGGCAAAG ACAAAAGTGTG TTTCCCTGT TAGTGTGTGG
TTGAGAAAGG AAACAAGAGT ATAAGGAGCT CTTAGATGAG
ACCTGTTCTT GTATCCTTGT GTCCATCTTC ATTCACTCAT
AGTCTCCAAA TGACTTTGTA TTTACATCAC TCCCAACAAA
TGTGATTCT TAATATTAG TTGACCATTA CTTTGTAA
30 TAAGACCTAC TTCAAAAATT TTATTCACTG TATTTCAGT
TGTGAGTCT TAATGAGGGA TACAGGAGGA AAAATCCCTA
CTAGAGTCCT GTGGGCTGAA ATATCAGACT GGGAAAAAAAT
GCAAAGCACA TTGGATCCTA CTTTCTCTA GATATTGAAC
CAGATCTCTG GCCCATCAGG CTTCTAAAT TCTTCAAAAG
35 AGCCACAACT TCCCCAGCTT CTCCAGCTCC CCTGCTCT
TCAATCCCTT GAGATATAGC AACTAACGAC GCTACTGGAA
GCCCCAGAGC AGAAAAGAAG CACATCTAA GATTCAAGGA
AAGACTAACT GTGAAAAGGA AGGCTGCTCT ATAACAAAGC
AGCATCAAGT CCCAAGTAAG GACAGTGAGA GAAAAGGGGG
40 AGAAGGAATG GAGCAAAGA GAACTGGCAA TAAGTAGGGG
AAGGAAGAAT TTCAATTGTC ATTGGGAGAG AGTTCTAAC
ACACTGAAGG CAACCTATT TCTACTGTT CTCTCTTGCC
AGGGTATTAG GAAGGACAGG AAAAGTAGGA GGAGGATCTG
GGGCATTGCC CTAGGAAATG AAAGAATTGT GTATAGAATG
45 GAAGGGGGAT CATCAAGGAC ATGTATCTCA AAATTTCTT
GAGATGCAGG TTAGTTGACC TTGCTGCAGT TCTCCCTCCC
ATTAATTCTAT TGGGATGGAA GCCAAAAATA AAAGAGGTGC
CTCTGAGGGAT TAGGGTTGAG CACTCAAGGG AAAGATGGAG
TAGAGGGCAA ATAGCAAAG TTGTTGCACT CCTGAAATT
50 TATTAACATT TCCGCAGAAG ATGAGTAGGG AGATGCTGCC
TTCCCTTTG AGATAGTGTAA GAAAAACACT AGATAGTGTG
AGAGGTTCCCT TTCTGTCCAT TGAAACAAGG CTAAGGATAC
TACCAACTAC TATCACCCTG ACCATTGTAC TGACAACAAT
TGAATGCAGT CTCCCTGCAG GGCAGATTAT GCCAGGCACT
55 TTACATTTGT TGATCCCATT TGACATTAC ACCAAAGCTC
TGAGTTCCAT TTACAGCTG AAGAAATTGA AGCTAGAGA
AATTAAGAAG CTGTTAAG TTTACACAGC TAGTAAGAGT
TTAAAAAATC TCTGTGCAGA AGTGTGGCT GGGTGTCTC
CCCACCACTA CCCTGTAAA CTTCCAGGAA GATTGGTTGA
60 AAGTCTGAAT AAAAGCTGTC CTTCTTACCA AATTCTTCC

CCCTCCTCAC TCTCACAAAGA AAACCAAAAG TTTCTCTTCA

SEQ ID NO:32

5 190428

Cluster name: G protein-coupled receptor Ls190428

SequenceID: AX100250

Sequence: GAGCAGAAAT TCGGCACGAG GAAAAATCTG AAATCTGAAA
 10 TGCTCCAAAAA TCCTAAACTT TTGAGTGCT GACATTATGC
 CACAAATGGAA AATTTCTATA CCTGACCTTA TGTGAGTTGC
 AGTCAAAACA CAGGTGCACA ACACCCAGTT CATGCAACAT
 CCCCAATGGG AAAAAGACC CCCCCAGCTC TCTTCTGCTG
 CAGTTTTCT GCTCACACCT GGATCCCCA TGCAATTCCA
 CAAAAAGTAA TAAATGGCA TGCCTGCAGG CTGGACACGC
 15 CAACAAACAGG TTCCCCACAA TGCCCCACAT GGGCGAAGAC
 CTGTGTGCAT TACTCATTGC ATTTTTTGC TTATTCTTG
 CTGTGTGGTA TAAATATATT GTTGAATAG TCAAAAAGAC
 CTAAAGATAC CCCTGTGAAT ATCAGTGATA AGAAAAAGAG
 GAAGCATTAA TGTTTATCTA TAGCACAGAA AGTCAAGTTG
 20 TTGGAGAAC TGGACAGTGG TGTAAGTGTG AAACATCTTA
 CAGAAGAGTA TGGTGTGGA ATGACCACCA TATATGACCT
 GAAGAACAG AAGGATAAAC TGTGAAGTT TTATGCTGAA
 AGTGTGAGC AGATATTAAAT GAAAAATAGA AAAACACTTC
 ATAAAGCTAA AAATGAAGAT CTTGATCGTG TATTGAAAGA
 25 GTGGATCCGT CAGCGTCGCA GTGAACACAT GCCACTTAAT
 GGTATGCTGA TCATGAAACA AGCAAAGATA TATCACAATG
 AACTAAAAAT TGAGGGGAAC TGTGAATATT CAAACAGGCTG
 GTTGCAGAAA TTAAAGAAAA GACATGGCAT TAAATTTTA
 AAGACTTGTG GCAATAAAGC ATCTGCTGGT CATGAAGCAA
 30 CAGAGAACIT TACTGGCAAT TTCAGTAATG ATGATGAACA
 AGATGGTAAC TTGAGGAT TCAGTATGTC AAGTGGAGAAA
 AAAATAATGT CTGACCTCCT TACATATACA AAAAATATAC
 ATCCAGAGAC TGTCAAG CTGGAAGAAG AGGATATCAA
 AGATGTTTT AACAGTAATA ATGAGGCTCC AGTGTGTCAT
 35 TCATTGTCCA ATGGTGAAGT AACAAAATG GTTCTGAATC
 AAGATGATCA TGATGATAAT GATAATGAAG ATGATGTTAA
 CACTGCAGAA AAAGTGCCTA TAGACGACAT GGTAAAAATG
 TGTGATGGGC TTATTAAGG ACTAGAGCAG CATGCATTCA
 TAACAGAGCA AGAAATCATG TCAGTTATA AAATCAAAGA
 40 GAGACTTCTA AGACAAAAG CATCATTAAAT GAGGCAGATG
 ACTCTGAAAG AAACATTAA AAAAGCCATC CAGAGGAATG
 CTTCTCCTC TCTACAGGAC CCACTTCTG GTCCCTCAAC
 TGCTTCTGAT GCTTCTTCTC ACCTAAAAT AAAATAAAAT
 ACAGTGTACA GTAACCTTT AGTCAAACAA GCATCATACT
 45 TGGAAACTGA AAGCC

SEQ ID NO:33

190437

Cluster name: G protein-coupled receptor C5L2

50 SequenceID: NM_018485

Sequence: CCTGTGTGCC ACGTGCTGGA CAAATCTAA CTCCTCAAGG
 ACTCCCAAA CCAGAGACAC CAGGAGCCTG AATGGGGAAAC
 GATTCTGTCA GCTACGAGTA TGGGGATTAC AGCGACCTCT
 CGGACCGCCC TGTGGACTGC CTGGATGGCG CCTGCCTGGC
 55 CATCGACCCG CTGCGCGTGG CCCCCCTCCC ACTGTATGCC

5 GCCATCTTCC TGGTGGGGGT GCCGGGCAAT GCCATGGTGG
 CCTGGGTGGC TGGGAAGGTG GCCCCCGGA GGGTGGGTGC
 CACCTGGTTG CTCCACCTGG CGTGGCGGA TTGCTGTGC
 TGTTTGTC TGCCCACCTCT GGCACTGCC ATTGCCGTG
 GAGGCCACTG GCCGTATGGT GCAGTGGCT GTCGGGCGCT
 GCCCTCCATC ATCCCTGCTGA CCATGTATGC CAGCGTCTG
 CTCCCTGGCAG CTCTCAGTGC CGACCTCTGC TTCTGGCTC
 TCGGGCCTGC CTGGTGGTCT ACGGTTCAAGC GGGCGTGCAG
 GGTGCAGGTG GCCTGTGGGG CAGCCTGGAC ACTGGCCTTG
 10 CTGCTCACCG TGCCCTCCGC CATCTACCGC CGGCTGCACC
 AGGAGCACTT CCCAGCCCGG CTGCAGTGTG TGGTGGACTA
 CGGCGGCTCC TCCAGCACCG AGAACATGCGGT GACTGCCATC
 CGGTTCTT TTGGCTTCCT GGGGCCCCCTG GTGGCCGTGG
 CCAGCTGCCA CAGTGCCTC CTGTGCTGGG CAGCCCGACG
 15 CTGCCGGCCG CTGGGCACAG CCATTGTGGT GGGGTTTTT
 GTCTGCTGGG CACCCCTACCA CCTGCTGGGG CTGGTGTCA
 CTGTGGCGGC CCCGAACCTCC GCACTCCTGG CCAGGGCCCT
 GCGGGCTGAA CCCCTCATCG TGGGCTTGC CCTCGCTCAC
 AGCTGCCCTCA ATCCCACATGCT CTTCTGTAT TTTGGGAGGG
 20 CTCAACTCCG CGGTCACTG CCAGCTGCCT GTCACTGGC
 CCTGAGGGAG TCCCAGGGCC AGGACGAAAG TGTGGACAGC
 AAGAAATCCA CCAGCCATGA CCTGGTCTCG GAGATGGAGG
 TGTAGGCTGG AGAGACATTG TGGGTGTGA TCTTCTTATC
 TCATITCACA AGACTGGCTT CAGGCATAGC TGGATCCAGG
 25 AGCTCAATGA TGCTTCATT TTATTCCTTC TTTCATCAA
 CAGATATCCA TCATGCACTT GCTATGTGCA AGGCCTTTT
 AGGCACCTAGA GATATAGCAG TGACCAAAAC AGACACAAAT
 CCTGCC

30 SEQ ID NO: 34

190701

Cluster name: C-C chemokine receptor 11

SequenceID: NM_016557

Sequence: CAAGACTGCT CCTCTCTGCC GACTACAACA GATTGGAGCC
 35 ATGGCTTGG AGCAGAACCA GTCAACAGAT TATTATTATG
 AGGAAAATGA AATGAATGGC ACTTATGACT ACAGTCAATA
 TGAACGTATC TGTATCAAAG AAGATGTCAAG AGAATTGTCA
 AAAGTTTCC TCCCTGTATT CCTCACAATA GTTTCGTCA
 TTGGACTITGC AGGCAATTCC ATGGTAGTGG CAATTATG
 40 CTATTACAAG AAACAGAGAA CCAAAACAGA TGTGTACATC
 CTGAATTGG CTGTAGCAGA TTACTCCTT CTATTCACTC
 TGCCCTTTG GGCTGTTAAT GCAGTTCATG GGTGGGTTTT
 AGGGAAATA ATGTCAAAA TAACCTCAGC CTGTACACCA
 CTAAACTTG TCTCTGGAAT GCAGTTCTG GCTTGTATCA
 45 GCATAGACAG ATATGTGGCA GTAACTAAAG TCCCCAGCCA
 ATCAGGAGTG GGAAAACCAT GCTGGATCAT CTGTTCTGT
 GTCTGGATGG CTGCCATCTT GCTGAGCATA CCCAGCTGG
 TTTTTATAC AGTAAATGAC AATGCTAGGT GCATTCCAT
 TTTCCCCCGC TACCTAGGAA CATCAATGAA AGCATTGATT
 50 CAAATGCTAG AGATCTGCAT TGGATTGTAA GTACCCCTTC
 TTATTATGGG GGTGTGCTAC TTATCACAG CAAGGACACT
 CATGAAGATG CCAAACATTA AAATATCTCG ACCCCTAAAAA
 GTTCTGCTCA CAGTCGTTAT AGTTTCATT GTCACTCAAC
 TGCCTTATAA CATTGTCAAG TTCTGCCAG CCATAGACAT
 55 CATCTACTCC CTGATCACCA GCTGCAACAT GAGCAAACGC
 ATGGACATCG CCATCCAAGT CACAGAAAGC ATCGCACTCT
 TTCACAGCTG CCTCAACCCA ATCCTTTATG TTTTTATGGG
 AGCATCTTC AAAAACTACG TTATGAAAGT GGCCAAGAAA

TATGGGTCTT GGAGAAAGACA GAGACAAAGT GTGGAGGAGT
 TTCTTGAGGGT CCTACAGAGC CAACCACTAC
 TTTTAGCAT TAAAGGTAAA ACTGCTCTGC CTTTGCTTG
 GATAACATATG AATGATGCTT TCCCCTCAAA TAAAACATCT
 5 GCATTATTCT GAAACTCAAA TCTCAGACGC CGTGGTTGCA
 ACTTATAATA AAGAATGGGT TGGGGGAAGG GGGAGAAATA
 AAAGCCAAGA AGAGGAAACA AGATAATAAA TGACAAAAAC
 ATGAAAATTA AAATGAACAA TATAGGAAAAA TAATTGTAAC
 AGGCATAAGT GAATAACACT CTGCTGTAAC GAAGAAGAGC
 10 TTTGTGGTGA TAATTTGTA TCTTGGITGC AGTGGTGCCTT
 ATACAAATCT ACACAAGTGA TAAAATGACA CAGAACTATA
 TACACACATT GTACCAATT CAATTCCCTG GTTTGACAT
 TATAGTATAA TTATGTAAGA TGGAACATT GGGGAAAAC
 15 GGGTGAAGGG TACCCAGGAC CACTCTGTAC CATCTTGTA
 ACTTCCTGTG AATTATAAT AATTCAAAA TAAAACAAGT
 TAAAAAAA CCCACTATGC TATAAGTTAG GCCATCTAAA
 ACAGATTATT AAAGAGGTT TC ATGTTAAAAG GCATTATAA
 TTATTTTAA TTATCTAAGT TTAATACAA GAACGATTTC
 CCTGCATAAT TTAGTACTT GAATAAGTAT GCAGCAGAAC
 20 TCCAACATAC TTTTCCCTG TTTTTTAA ATTGTAAGT

SEQ ID NO: 35

190705

25 Cluster name: G-protein coupled receptor SALPR
 SequenceID: NM_016568
 Sequence: GATTTGGGGA GTTATGCGCC AGTGCCAG TGACCGCGGG
 ACACGGAGAG GGGAAAGTCTG CGTTGTACAT AAGGACCTAG
 30 GGAAGTCCGAG CTTGGCCTGA GAACCTTGG ACGCCGAGTG
 CTTGCCCTAAC GGGCTGCACT CCTCAACTCT GCTCAAAGC
 AGCCGCTGAG CTCAACTCCT GCGTCCAGGG CGTCGCTGC
 GCGCCAGGAC GCGCTTAGTA CCCAGTTCTT GGGCTCTC
 TTCAGTAGCT GCTTGAAG CTCCACGCCA CGTCCCGCAG
 GCTAGCCTGG CAACAAAAGT GGGTAAACC GTGTTATCTT
 35 AGGTCTTGTC CCCCCAGAAC TGACCTAGAG GTACCTGCGC
 ATGCAGATGG CCGATGCAGC CACGATAGCC ACCATGAATA
 AGGCAGCAGG CGGGGACAAG CTAGCAGAAC TCTTCAGTCT
 GGTCCCGGAC CTTCTGGAGG CGGCCAACAC GAGTGGTAAC
 GCGTCGCTGC AGCTCCGGA CTTGTGGTGG GAGCTGGGC
 40 TGGAGTTGCC GGACGGCGCG CGGCCAGGAC ATCCCCCGGG
 CAGCGCGGG GCAGAGAGCG CGGACACAGA GGCCCGGGTG
 CGGATTCTCA TCAGCGTGGT GTACTGGGTG GTGTGCGCCC
 TGGGGTTGGC GGGCAACCTG CTGGTTCTCT ACCTGATGAA
 45 GAGCATGCAG GGCTGGCGCA AGTCCTCTAT CAAACCTCTC
 GTCACCAACC TGGCGCTGAC GGACTTCAG TTTGTGCTCA
 CCCTGCCCTT CTGGCGGTG GAGAACGCTC TTGACTTCAA
 ATGGCCCTTC GGCAAGGCCA TGTGTAAGAT CGTGTCCATG
 GTGACGTCCA TGAACATGTA CGCCAGCGTG TTCTCCTCA
 CTGCCATGAG TGTGACGCGC TACCATTCGG TGGCCTCGGC
 50 TCTGAAGAGC CACCGGACCC GAGGACACGG CGGGGGCGAC
 TGCTGCGGCC GGAGCCTGGG GGACAGCTGC TGCTTCTCGG
 CCAAGGGCCT GTGTGTTGGG ATCTGGGCTT TGGCCCGCGCT
 GGCCTCGCTG CCCAGTGCCA TTTTCTCCAC CACGGTCAAG
 GTGATGGGCG AGGAGCTGTG CCTGGTGCCTT TTCCCGGACA
 55 AGTTGCTGGG CGCGACAGG CAGTTCTGGC TGGGCTCTA
 CCACTCGCAG AAGGTGCTGT TGGGCTCTCGT GCTGCCGCTG
 GGCATCATTAA TCTTGTGCTA CCTGCTGCTG GTGCCGCTCA

TCGCCGACCG CCGCGCGGCG GGGACCAAAG GAGGGGCCGC
 GGTAGCCGGA GGACGCCGA CGGGAGCCAG CGCCCGGAGA
 CTGTCGAAGG TCACCAAATC AGTGACCATC GTTGTCTGT
 CCTCTTCT GTTGTGGCTG CCCAACCAAGG CGCTCACCA
 5 CTGGAGCATC CTATCAAGT TCAACCGGGT GCCCTTCAGC
 CAGGAGTATT TCCTGTCCA GGTATACGCG TTCCCTGTGA
 GCGTGTGCCT AGCGCACTCC AACAGCTGCC TCAACCCCCT
 CCTCTACTGC CTCGTGCGCC GCGAGTTCCG CAAGGCGCTC
 AAGAGCCTGC TGTGGCGCAT CGCGTCTCCT TCGATCACCA
 10 GCATGCGCCC CTTACCGCC ACTACCAAGC CGGAGCACGA
 GGATCAGGGG CTGCAGGCC CGGGGCCGCC CCACGCGGCC
 GCGGAGGCCGG ACCTGCTCTA CTACCCACCT GGCGTCGTGG
 TCTACAGCGG GGGCGCTAC GACCTGCTGC CCAGCAGCTC

15

SEQ ID NO: 36

190711

Cluster name: G protein-coupled receptor GPR85

SequenceID: NM_018970

20 Sequence: GGCACGAGGA TTTTACTGCT GTCTCAAGAT CAGATTATTA
 CTGTAGAGAA GATTTTATT TTTTGTTC TAAACAGATT
 ATTATAAAGC AAAAAGCATG CAGAAAAAAGA AGCAGACGTT
 TTACATTGGG ATTAATGAA AGCGTGTCTG CTAGTTTGG
 GTAGGAGAAC TGGGAAGTTG TTGCTAAAAA TTTTATATCA
 25 CCTCCACAAA CAAAACCTTT CGGAAATGGT AAAATAAGAA
 AATGCATGAT TCTAGAGGCA TTCCTAACGA CCCACGTGTC
 AGGCCTTGTG GTGTCGTGG TATCATCCGA CGGTTGGAC
 TGGTTAGGGC TTACTGAGAG CTCCATTCT GGAAAGCCTT
 ACAAGACTGA GGAATATCAG ACTGCGAATC ACCGGGAACG
 30 GTTCCTTGCA AGCACAGAAG CAATCTCTCT CCCCATCTTC
 GCATATTCTG ATGGCAAAAC AAGTGGAAAG AAAGAGGAAG
 CATGACTGCA GATCAGATCA GTTCTCTTG TGGATTATAT
 TTTCAGTAAA ATGTATGGAT CTATCTTTC CTTGTTCTTA
 TATCTAGATC ATGAGACTTG ACTGAGGCTG TATCCTTATC
 35 CTCCATCCAT CTATGGCGAA CTATAGCCAT GCAGCTGACA
 ACATTTGCA AAATCTCTCG CCTCTAACAG CCTTCTGAA
 ACTGACTTCC TTGGGTTTC TAATAGGAGT CAGCGTGGTG
 GGCAACCTCC TGATCTCCAT TTTGCTAGTG AAAGATAAGA
 CTTGCGATAG AGCACCTITAC TACTTCTGT TGGATCTTTG
 40 CTGTTAGAT ATCCTCAGAT CTGCAATTG TTTCCCATTT
 GTGTTCAACT CTGTCAAAAA TGGCTCTACC TGGACITATG
 GGACTCTGAC TTGCAAAGTG ATTGCCTTC TGGGGGTTTT
 GTCCCTGTTT CACACTGCTT ICATGCTCTT CTGCATCAGT
 GTCACCAGAT ACTTAGCTAT CGCCCATCAC CGCTTCTATA
 45 CAAAGAGGCT GACCTTTGG ACGTGTCTGG CTGTGATCTG
 TATGGTGTGG ACTCTGTCTG TGGCCATGGC ATTTCCCCCG
 GTTTTAGACG TGGGCACTTA CTCATTCTT AGGGAGGAAG
 ATCAATGCAC CTTCCAACAC CGCTCCTTC GGGCTAATGA
 TTCTTAGGA TTATGCTGC TTCTGCTCT CATCCTCTTA
 50 GCCACACAGC TTGTCACCT CAAGCTGATA TTTTCGTCC
 ACGATCGAAG AAAAATGAAG CCAGTCCAGT TTGTAGCAGC
 AGTCAGCCAG AACTGGACTT TTCAATGGTCC TGGAGCCAGT
 GGCCAGGAG CTGCCAATTG GCTAGCAGGA TTTGGAAGGG
 GTCCCACACC ACCCACCTTG CTGGGCATCA GGCAAAATGC
 55 AAACACCCA GGCAGAAGAA GGCTATTGGT TTAGACGAG
 TTCAAAATGG AGAAAAGAAT CAGCAGAATG TTCTATATAA
 TGACTTTCT GTTCTAACCT TTGTGGGGCC CCTACCTGGT
 GCCTGTTAT TGGAGAGTTT TTGCAAGAGG GCCTGTAGTA

CCAGGGGGAT TTCTAACAGC TGCTGTCTGG ATGAGTTTG
CCCAAGCAGG AATCAATCCT TTTGTCTGCA TTTTCTCAA
CAGGGAGCTG AGGCAGCTGTT TCAGCACAAC CCTTCTTAC
TGCAGAAAAT CCAGGTTACC AAGGGAACCT TACTGTGTTA
5 TATGAGGGAG CATCTGAAA TCTTTAGCCT TGTGAAAATC
AACCTCTCT GCTGAGCAAT TGTGGCCCAT AGCCATATT
TGAGAAGAAA TTCAAGAATG GAATCAGCAG TTTAAGGAT
TTGGGCAACA TTCTGCAGTC TTTGCAATAG TTCAACCTATA
ATCCTATTAA AAATCTCAGA GTGATCCTGC TGACTGCCAG
10 CAAAGGTTG TAATTAAGAA GGGACTGAAC CACTGCCATA
AGTTTCTTA TGTGGTCAAA AACTAGATAA TGAAAGTAGC
AGGTGCTAAG TATCAGTGT AAATGCTTG TATGTCACTA
CATATGAAAA AACATCAAAA ACAATTAGC ATTGGACATC
TTAATAAATT AAGTTGACAT GAGGTAATG TGTGATAAA
15 AACTAATTAA AGAAGTTGA AGACTTTAAA ACATTTATA
CTACTATTGT TTTGCAAAGA CTAAAATATT TGGGGACTTA
AAGTACTGTA ATCCACTAAA GACGTGCCAA TGAATTATTG
GAATATCACA CTTTAAAC CGCCTTGAA GTTCTGGGA
GCATTCCAAA GCAGTATATT GGTTCCAATT AGAGTTACT
20 TTTTTGTAT TAATACATTG CTATTCTAA ATACCACTTT
CCTCATCTAC TAGTAAGATT GCTAGCATTG AACTGTATTA
TGTGGTTTT GTTGATTGG TATAAGTTT TTCCAATTCA
TTTATATTAA ACAAAATGCTA GATATTGGTC TGGGAGGCAA
CATTAATGGT ACCAGCCTGT CACAACGTGAG CAGTTCTAAT
25 AATGCAGAAAT AAATACATGT TGCCCTAAAG GGTATCTAG
TATCCTTCAT CTTATTAGC ACTGGAGCAA ATAGCCAAGG
GAAATCAAAT CAGTAACCTGG TCATGGTCAT GCATCTAAA
GTGCATGGAA GATCATTAT TACTTTTCC TTTTTCTC
ACATGGTTTG AAACCTAAAG TGACACATCAC TGAATAATG
30 AGATTTCCTT CTACGGTGTG CTACCCCTTC TAAACTGTTC
TAAGAAGCAG GCAGTTGATG TATGTTATA TTTTAAGTCA
GCTGTCAAGG GGAGACCAAC GCCTTAGTAT GACATCCTGC
ACAATTGTG AAGCATTAT TCTACTGAAG GCACAGTC
GTTTATACTT TCTGCACATT CAGTGTATTG GTAATTAA
35 TTATTCAGT TTAAACTTGT GAAAGCTTAT ATTATGATT
CTGGTATTAT AGAAATACAT TAGAGTCTGT GAGTCTCATT
CTTTAAGATA CAGATGTGTG AACTCAATA TAAAGTTGCA
TTGCCAAAA TTACCCGTG TAGCCTGTTA ATTTCTTGA
AATAAGTTT ACATTTGG CACATAACAA CGTTTTTT
40 AATTGGGAG GCAAGCACAA ACTAGGAAGA CTAGCTTAT
TATGGTTTG CTTTTGATT CTGTAGCTA CTATATTCCA
GAETGGAAAT GTATGAATGA TAATCAACAT AATGCTGATA
AACTGACATA ATATTATCTG TAAAAGCATT ATTTGGTAGT
TTATTATAAT CATCCCTCTA TTATTCTTAA ATGCCAGTAG
45 TATTTAGAGA TGTGTACCTG CTTAGTTAT TGGCTCAGAA
TTTAATATA AACATCACAC TTAAATTGG AGCATAGTAC
CATAGAAATT TGGGGTTCTA AATATACAAC TTGTAAGAAG
AATGGTTAC ACTAACATTA TGACAAAAT AGAAAAAGTT
ATTATTITG TTGCTTCT GTGTTTGTTGTTGTTGTTG
50 GTTTTGTTGTA AGTTTATTAT TTTTTGGTA TTGATAATT
AAGATTAGGA ATCTAATAAC ACAGAATTCC ATATTGCTAT
AGTACTTCTG TAAAGAGAAT ATCAATATAA ATAAGGAAAA
TAAATCAATG AAATGTTCA ATGGTTAAAA AAAAAAAA AAAAA
55 SEQ ID NO:37

190774

Cluster name: Histamine H4 receptor

SequenceID: NM_021624

Sequence: GAATTGTCTG GCTGGATTAA TTTCATAATT TGACCTTCTT
 CATCATTTGA TGTGATGCCA GATACTAATA GCACAATCAA
 5 TTTATCACTA AGCACTCGTG TTACTTTAGC ATTTTTATG
 TCCTTAGTAG CTTTGCTAT AATGCTAGGA AATGCTTGG
 TCATTTAGC TTTGTGGTG GACAAAAAACC TTAGACATCG
 AAGTAGTTAT TTTTTCTTA ACTTGGCCAT CTCTGACTTC
 10 TTGTGGGTG TGATCTCCAT CCCTTGTAC ATCCCTCACA
 CGCTGTTCGA ATGGGATTAA GGAAAGGAAA TCTGTGTATT
 TTGGCTCACT ACTGACTATC TGTTATGTAC AGCATCTGTA
 TATAACATTTG TCCTCATCAG CTATGATCGA TACCTGTCAG
 TCTCAAATGC TGTGCTTAT AGAACTCAAC ATACTGGGGT
 15 CTTGAAGATT GTACTCTGA TGGTGGTCGT TTGGGTGCTG
 GCCTTCTAG TGAATGGGCC AATGATCTA GTTTCAGAGT
 CTTGGAAGGA TGAAGGTAGT GAATGTGAAC CTGGATTTTT
 TTCGGAATGG TACATCCTTG CCATCACATC ATTCTTGGAA
 20 TTCTGATCC CAGTCATCTT AGTCGCTTAT TTCAACATGA
 ATATTATTC GAGCCTGTGG AAGCGTGTAC GTCTCAGTAG
 GTGCCAAAGC CATCCGGAC TGACTGCTGT CTCTTCCAAC
 ATCTGTGGAC ACTCATTCAAG AGGTAGACTA TCTICAAGGA
 25 GATCTCTTC TGATCGACA GAAGTTCTG CATCCTTCA
 TTCAAGAGAGA CGGAGGAGAA AGAGTAGTCI CATGTTTTC
 TCAAGAACCA AGATGAATAG CAATACAATT GCTTCCAAA
 TGGGTTCTT CTCCCAATCA GATTCTGTAG CTCTTCACCA
 AAGGGAACAT GTTGAACACTGC TTAGAGCCAG GAGATTAGCC
 30 AAGTCACTGG CCATTCTCTT AGGGTTTT GCTGTTGCT
 GGGCTCCATA TTCTCTGTTC ACAATTGTCC TTTCATTTTA
 TTCTCAGCA ACAGGTCTTA AATCAGTTTG GTATAGAATT
 GCATTTGGC TTCAAGTGGTT CAATTCCCTT GTCAATCCTC
 TTTGTATCC ATTGTGTAC AAGCGCTTTC AAAAGGCTTT
 CTTGAAAATA TTTGTATAA AAAAGCAACC TCTACCATCA
 CAACACAGTC GGTAGTATC TTCTTAAAGA CAATTITCTC
 ACCTCTGTAA ATTTAGTCT CAATC

SEQ ID NO:38

35 191168

Cluster name: P2Y12 platelet ADP receptor

SequenceID: NM_022788

Sequence: GGCTGCAATA ACTACTACTT ACTGGATACA TTCAAACCCCT
 40 CCAGAACAA CAGTTATCAG GTAACCAACA AGAAATGCAA
 GCCGTCGACA ACCTCACCTC TGCGCCTGGG AACACCAGTC
 TGTGCACCAAG AGACTACAAA ATCACCCAGG TCCTCTTCCC
 ACTGCTCTAC ACTGCTCTGT TTTTGTGG ACTTATCACA
 AATGGCCTGG CGATGAGGAT TTTCTTCAA ATCCGGAGTA
 45 AATCAAACCT TATTATTTT CTTAAGAACAA CAGTCATTC
 TGATCTCTC ATGATTCTGA CTTTCCATT CAAAATTCTT
 AGTGATGCCA AACTGGGAAC AGGACCACTG AGAACCTTGT
 TGTGTCAAGT TACCTCCGTC ATATTTATT TCACAAATGTA
 TATCAGTATT TCATTCCTGG GACTGATAAC TATCGATCGC
 TACCAGAAGA CCACCAGGCC ATTTAAACAA TCCAACCCCCA
 50 AAAATCTCTT GGGGGCTAAG ATTCTCTCTG TTGTCACTCG
 GGCATTCCATG TTCTTACTCT CTTTGCTAA CATGATTCTG
 ACCAACAGGC AGCCGAGAGA CAAGAAATGTG AAGAAATGCT
 CTTTCCCTAA ATCAGAGTTC GGTCTAGTCT GGCATGAAAT
 AGTAAATTAC ATCTGTCAAG TCATTTCTG GATTAATTTC
 55 TTAATTGTAA TTGTATGTAA TACACTCATT ACAAAAGAAC
 TGTACCGGGTC ATACGTAAGA ACGAGGGGTG TAGGTAAAGT
 CCCCAGGAAA AAGGTGAACG TCAAAGTTT CATTATCATT
 GCTGTATTCT TTATTGTGTGTT TTGTCTTTC CATTITGCC

GAATTCCCTTA CACCCGTGAGC CAAACCCGGG ATGTCTTTGA
CTGCACTGCT GAAAATACTC TGTTCTATGT GAAAGAGAGC
ACTCTGTGGT TAACCTCCTT AAATGCATGC CTGGATCCGT
TCATCTATT TTTCCCTTGC AAGTCCTTCA GAAATCCTT
5 GATAAGTATG CTGAAGTGCC CCAATTCTGC AACATCTCG
TCCCAGGACA ATAGGAAAAA AGAACAGGAT GGTGGTGACC
CAAATGAAGA GACTCCAATG TAAACAAATT AACTAAGGAA
ATATTCAT CTCTTGTGT TCAGAACTCG TTAAAGCAAA
10 GCGCTAAGTA AAAATATTAA CTGACGAAGA AGCAACTAAG
TTAATAATAA TGACTCTAAA GAAACAGAAG ATTACAAAAG
CAATTTCAT TTACCTTCC AGTATGAAAAA GCTATCTAA
AATATAGAAA ACTAATCTAA ACTGTAGCTG TATTAGCAGC
AAAACAAACG AC

15 SEQ ID NO:39

191218

Cluster name: G protein-coupled receptor Ls191218

SequenceID: AX099247

Sequence: TTAATCTCTT CAAGCCTCTG ATTTCTCTC CTGTAAAACA
20 GGGGCGGTAA TTACACATA ACAGGCTGGT CATGAAAATC
AGTGAACATG CAGCAGGTGC TCAAGTCTTG TTTTGTTTC
CAGGGGCACC AGTGGAGGTT TTCTGAGCAT GGATCCAACC
ACCCCGGCCT GGGGAACAGA AAGTACAACA GTGAATGGAA
ATGACCAAGC CCTTCTTCTG CTTGTGGCA AGGAGACCC
25 GATCCCGGTC TTCTGATCC TTTCATTCG CCTGGTCGGG
CTGGTAGGAA ACGGGTTTGT GCTCTGGCTC CTGGGCTTCC
GCATGCGCAG GAACGCCTTC TCTGTCTACG TCCTCAGCCT
GCCCGGGGCC GACTTCCTCT TCCTCTGCTT CCAGATTATA
AATTGCCTGG TGTACCTCAG TAACTCTTC TGTTCCATCT
30 CCATCAATT CCCTAGCTTC TTCACCACTG TGATGACCTG
TGCCTACCTT GCAGGCCTGA GCATGCTGAG CACCGTCAGC
ACCGAGCGCT GCCTGTCCGT CCTGTGGCCC ATCTGGTATC
GCTGCCGCCG CCCCAGACAC CTGTCAGCGG TCGTGTGTGT
CCTGCTCTGG GCCCTGTCCC TACTGCTGAG CATCTGGAA
35 GGGAAAGTTCT GTGGCTTCTT ATTAGTGAT GGTGACTCTG
GTTGGTGTCA GACATTGAT TTCATCACTG CAGCGTGGCT
GATTTTTTA TTCATGGTTC TCTGTGGGTG CAGTCTGGCC
CTGCTGGTCA GGATCCTCTG TGGCTCCAGG GGTCTGCCAC
TGACCAGGCT GTACCTGACC ATCCTGCTCA CAGTGCTGGT
40 GTTCCTCCTC TGCAGGCCTGC CCTTGCGAT TCAGTGGTTC
CTAATATTAT GGATCTGGAA GGATTCTGAT GTCTTATTIT
GTCATATTCA TCCAGTTCA GTTGTCTGT CATCTCTAA
CAGCAGTGCC AACCCCATCA TTAACTCTT CGTGGGCTCT
45 TTTAGGAAGC AGTGGCGGCT GCAGCAGCCG ATCCTCAAGC
TGGCTCTCCA GAGGGCTCTG CAGGACATTG CTGAGGTGGA
TCACAGTGAA GGATGCTTCC GTCAGGGCAC CCCGGAGATG
TCGAGAAGCA GTCTGGTGTGAGATGGACA GCCTCTACTT
CCATCAGATA TATGTG

50 SEQ ID NO:40

189884

Cluster name: G protein-coupled receptor LS189884

SequenceID: ENSMDNA108574

Sequence: ATGCTGGCAG CTGCCCTTGC AGACTCTAAC TCCAGCAGCA TGAATGTGTC
55 CTTTGCTCAC CTCCACCTTG CCGGAGGGTA CCTGCCCTCT GATTCCCAGG ACTGGAGAAC

CATCATCCCG GCTCTCTGG TGGCTGTCIG CCTGGTGGGC TTCGTGGAA ACCTGTGTGT
 GATTGGCAGTC CTCCCTCACCA ATGCTTGGAA AGGAAAGCCA TCCATGATCC ACTCCCTGAT
 TCTGAATCTC AGCCTGGCTG ATCTCTCCCT CCTGCTGTT TCTGCACCTA TCCGAGCTAC
 GGCCTACTCC AAAAGTGTGTT GGGATCTAGG CTGGTTGTC TGCAAGTCCT CTGACTGGTT
 5 TATCCACACCA TGCATGGCAG CCAAGAGCCT GACAATCGTT GTGGTGGCCA AAGTATGCTT
 CATGTATGCA AGTGACCCAG CCAAGCAAGT GAGTATCCAC AACTACACCA TCTGGTCAGT
 GCTGGTGGCC ATCTGGACTG TGGCTAGCCT GTTACCCCTG CCGGAATGGT TCTTAGCAC
 CATCAGGCAT CATGAAGGTG TGGAAATGTG CCTCGTGGAT GTACCAGCTG TGGCTGAAGA
 GTTATGTGCG ATGTTGGTA AGCTCTACCC ACTCCTGGCA TTTGGCCTTC CATTATTTT
 10 TGCCAGCTT TATTCTGGA GAGCTTATGA CCAATGTAAA AAACGAGGAA CTAAGACTCA
 AAATCTTAGA AACCAAGATAC GCTCAAAGCA AGTCACAGTG ATGCTGCTGA GCATTGCCAT
 CATCTCTGCT CTCTTGTGGC TCCCCGAATG GGTAGCTTGG CTGTGGGTAT GGCAATCTGAA
 GGCTGCAGGC CGGGCCCCAC CACAAGGTTT CATAGCCCTG TCTCAAGTCT TGATGTITC
 15 CATCTCTICA GCAAATCTC TCATTTTCT TGTGATGTCG GAAGAGTCA GGGAAAGGCTT
 GAAAGGTGTA TGGAAATGGA TGATAACCAA AAAACCTCCA ACTGTCTCAG AGTCTCAGGA
 AACACCAGCT GGCAACTCAG AGGGTCTCC TGACAAGGTT CCATCTCCAG AATCCCCAGC
 ATCCATACCA GAAAAAGAGA AACCCAGCTC TCCCTCCTCT GGCAAAGGGA AAACGTGAGAA
 GGCAGAGATT CCCATCCTTC CTGACGTAGA GCAGTTTG CATGAGAGGG ACACAGTCCC
 20 TTCTGTACAG GACAATGACC CTATCCCTG GGAACATGAA GATCAAGAGA CAGGGGAAGG
 TGTTAAATAG

SEQ ID NO:41

168928

25 Cluster name: G protein-coupled receptor Ls168928

SequenceID: AW973537

Sequence: AGTAGTAATC TCATCTTGTG CACTGTGGGG TCTTCTAATG
 TGACCCTGAG CAATCTCTG CATACCAAGTA AAGACTGTT
 ACTTTTCCAC CATGAACCTCC ATCATCAGAA GACTGTTCT
 30 TACTCTGTTT CTTACTCCAG ATATGTTTT CTTATAGGAA
 CAATGCTGCT TTCAAGTGCA TACAGAGTGG TCCTTTGTT
 CAGGCACCCAG AAGAAAATTCT GATACTTTCA CAGCACCAGC
 CTTTCCCCAA GACCTTCCCC AGAGAAAAGT GCCACTCAGA
 CCATCCTGCT GCTAGTGAGT TTCTTGTTGG TCATCTACTG
 35 GGTGATTTTC ATCATCTCAT GCACCTCAAC CTTGCTATGG
 GCATATGACC CTGTTGTCCT GGGTGTCCAG AGGCTTGTCA
 GTCTTTGTT GCTACTCAGA TCTGATAAAA GGATAATCAT
 TGTGACACAA ACTGTGAGAC AGATGGTAA CAAGTTATT
 40 TTATTGAAAA TAGATTATTC TGTCACCAGT TAAATTACAT
 AAGTAGTACA GAACTTGCTA TTTAATTAAAC TTAAATGGTT
 GGATTACAC TTTCAATATG

SEQ ID NO:42

189890

45 Cluster name: G protein-coupled receptor Ls189890

SequenceID: ENSMDNA279706

Sequence: CTTCCTCATC AGACTGTTGC CTGGCTACAC GGCTGGGC
 AGCGCCAACA GGAAGTCCTT AAAGGCAGGT ATTATTC
 50 AGTGTATGGT CAGGCTCAAG CTGCCATTCA GCAACTCGTG
 GGCTTTGGGA CCCAGCACCG AGGGGTTATA TGTGAAGGAG
 GGCCCCCGCC AGGAGTCTGA AGTAAAATG GTAGCAGTCA
 CAGACAATGA CGGTGGCAGC AGGGGTTAG GCAATGACGG
 TGGCCATGCT GTTGATGCTG TCATCTACAC TGCTGATCTT TGA

SEQ ID NO: 43

189893

Cluster name: G protein-coupled receptor Ls189893

SequenceID: AI285887

5 Sequence: TTTGTGTACA AGAATTTAT GTACTTTAAC TACTGTGGCA
CAAGTGACAT GGCCAAAATG GACCTTCCT CCAACACACT
GGTGTGTGG CGTCTGCTGC CTGGTGCCAC CTATAACAAC
CGCTTTCT ATGCTGGTGT GCCCTGGAAG GACTTAGATT
TTGCTGGTGA TGAGAAGGGG CTGTGGGTC TCTATGCCAC
10 TGAGGAGAGC AAGGGCAACC TGGTTGTGAG TCGTCTCAAC
GCTAGCACCC TAGAAGTGGA GAAAACCTGG CGTACCAAGCC
AGTACAAGCC AGCCCTGTCA GGGGCCTTCA TGGCCTGTGG
GGTGTCTAT GCCTTACACT CACTGAACAC CCACCAAGAG
GAGATCTTCT ATGCTTTGA CACCACCACC GGG

15

INTERNATIONAL SEARCH REPORT

Inte al application No.

PCT/US01/15332

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07K 14/705, 16/28; C12N 15/12

US CL : 435/7.1, 69.1, 252.3, 320.1; 530/350, 388.22; 536/23.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 69.1, 252.3, 320.1; 530/350, 388.22; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Geneseq, Issued Patents, EST
searched SEQ ID NO:3**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| A,P | WHITE et al. (The ADHR Consortium), Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF23. Nature Genetics. November 2000. Vol. 26. No.3. pages 345-348. see entire document. | 1-10, 14-18 |
| A,P | WO 01/04292 A1 (MERCK PATENT GMBH) 18 January 2001. SEQ ID NO:1. | 1-10, 14-18 |

 Further documents are listed in the continuation of Box C. See patent family annex.

| | | |
|--|-----|---|
| Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" | | document defining the general state of the art which is not considered to be of particular relevance |
| "E" | "X" | earlier document published on or after the international filing date |
| "L" | | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) |
| "O" | "Y" | document referring to an oral disclosure, use, exhibition or other means |
| "P" | "Z" | document published prior to the international filing date but later than the priority date claimed |

Date of the actual completion of the international search

27 SEPTEMBER 2001

Date of mailing of the international search report

25 OCT 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

JOHN ULM

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

| | |
|----------------|--------------------|
| Inte | al application No. |
| PCT/US01/15352 | |

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-10 and 14 to 18 in so far as they relate to SEQ ID NO:s.

Remark on Protest

| |
|--------------------------|
| <input type="checkbox"/> |
| <input type="checkbox"/> |

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

| | |
|----------------|-------------------|
| Inte | 1 application No. |
| PCT/US01/15392 | |

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The different species consist of the 48 nucleotide sequences listed in Table 1 of the instant description and 48 antibodies which bind to 48 different polypeptides.

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I-XLVIII, claims 1 to 10 and 14 to 18, which are drawn to an isolated polynucleotide encoding any one of 48 different polypeptides, an isolated polypeptide encoded by that nucleic acid and methods of use.

Group II, XLIX-XCVI, claims 11 to 19, drawn to an antibody which binds to any one of 48 different polypeptides.

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the nucleic acids and proteins of invention I do not share a common utility with the antibodies of invention II and each of these inventions can be made and used without the other.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: The 48 nucleic acids listed in Table 1 of the instant description lack a common utility which is based upon a special technical feature which is common to all of those nucleic acids and which is lacking from the prior art.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
15 November 2001 (15.11.2001)

PCT

(10) International Publication Number
WO 01/85791 A1

(51) International Patent Classification⁷: C07K 14/705. (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(21) International Application Number: PCT/US01/15332

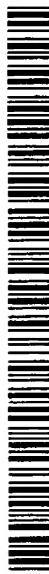
(22) International Filing Date: 11 May 2001 (11.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/203,217 11 May 2000 (11.05.2000) US
60/205,945 18 May 2000 (18.05.2000) US

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).



(71) Applicant (*for all designated States except US*): LIFESPAN BIOSCIENCES, INC. [US/US]; 700 Blanchard Street, Seattle, WA 98121 (US).

Published:

— with international search report

(48) Date of publication of this corrected version:
7 February 2002

(15) Information about Correction:

see PCT Gazette No. 06/2002 of 7 February 2002, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and
(73) Inventors/Applicants (*for US only*): BROWN, Joseph, P. [GB/US]; 411 West Prospect Street, Seattle, WA 98119 (US). MILLER, Margaret [US/US]; 7155 S.E. Maker Street, Mercer Island, WA 98040 (US). BURMER, Glenna [US/US]; 7516 55th Place, NE, Seattle, WA 98115 (US). FABRE-SUVER, Christine [FR/US]; 1212 NE 63rd Street, Seattle, WA 98115 (US). PRITCHARD, David [US/US]; 3429 Burke Avenue, N., Seattle, WA 98103 (US).

A1
WO 01/85791 A1

(54) Title: NUCLEIC ACID SEQUENCES FOR NOVEL GPCRS

(57) Abstract: The present invention is directed to isolated nucleic acid encoding G protein-coupled receptors and the proteins encoded thereby. Included are methods of using the disclosed nucleic acid and protein.

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)